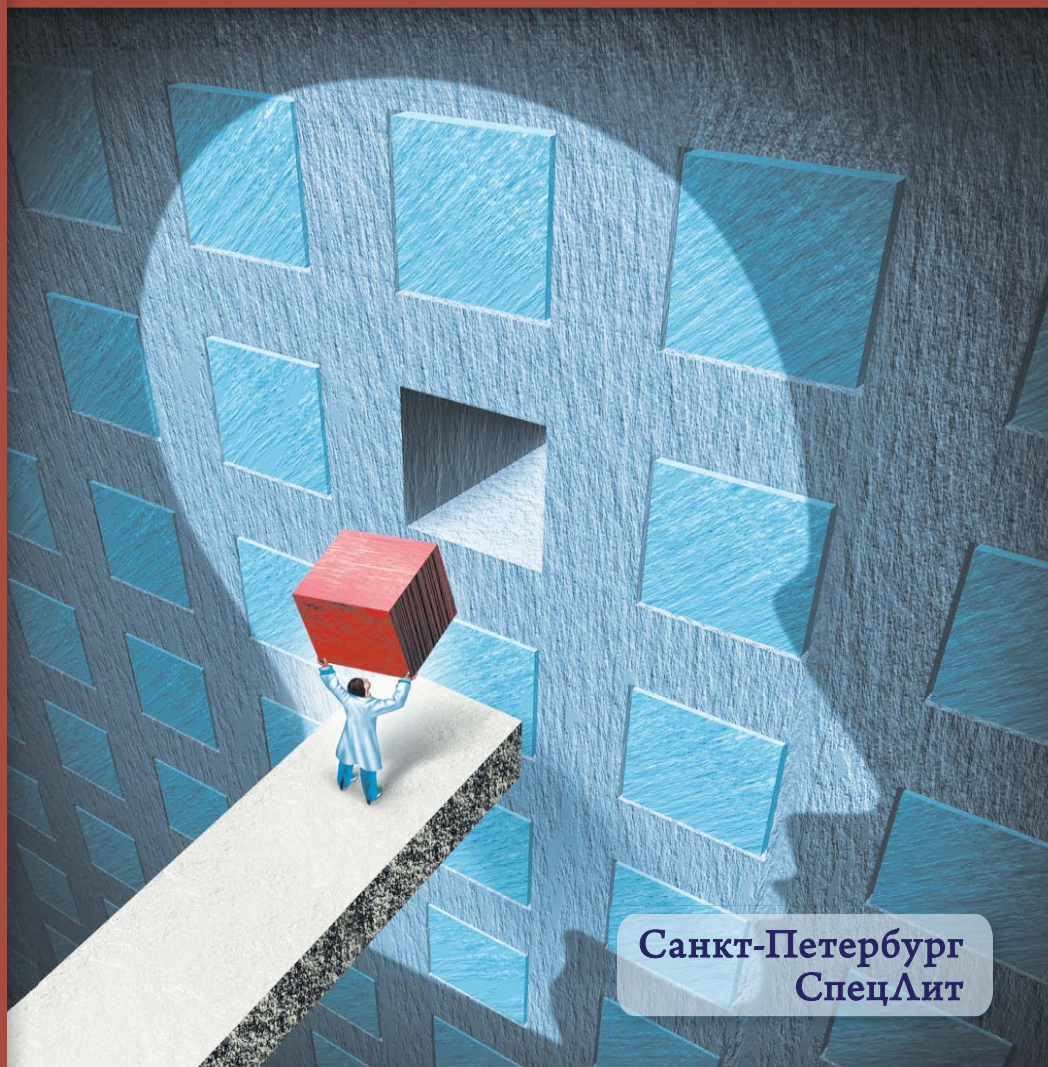


И. А. Кельмансон

Методология исследования в клинической психологии

Учебное пособие



Санкт-Петербург
СпецЛит

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ В КЛИНИЧЕСКОЙ ПСИХОЛОГИИ

Учебное пособие

Санкт-Петербург
СпецЛит
2017

УДК 159.9
К-34

Рецензенты:

Вартамян И. А. — д-р биол. наук, канд. мед. наук, профессор. Главный научный сотрудник Института эволюционной физиологии и биохимии им. И. М. Сеченова РАН;

Шпиленя Л. С. — д-р мед. наук, профессор. Профессор кафедры клинической психологии ЧОУВО «Институт специальной педагогики и психологии».

Кельмансон И. А.

К-34 Методология исследования в клинической психологии : учеб. пособие. — Санкт-Петербург : СпецЛит, 2017. — 328 с.
ISBN 978-5-299-00858-6

Пособие посвящено методологии исследования в клинической психологии. Рассматривается подготовительный этап исследования, особое внимание уделяется подходам к поиску, отбору и систематическому анализу литературных источников. Подробно обсуждаются возможные виды дизайна исследования, в том числе описательные и аналитические исследования и их разновидности. Описан процесс измерения, его надежность и валидность. Рассматриваются вопросы валидности исследования, подходы к снижению угроз валидности. Приводятся сведения о подходах к анализу данных и об оценке выраженности эффекта, оптимизации письменного представления сведений об исследовании в научных публикациях.

Пособие рассчитано на студентов, обучающихся по специальности «Клиническая психология», аспирантов, научных сотрудников и практических клинических психологов.

УДК 159.9

Подписано в печать 20.03.2017. Формат 60 × 88^{1/16}.
Печ. л. 20,5. Тираж 1000 экз. Заказ №
ООО «Издательство „СпецЛит“».
190103, Санкт-Петербург, 10-я Красноармейская ул., 15
Тел.: (812) 495-36-09, 495-36-12
<http://www.speclit.spb.ru>.
Отпечатано в ГП ПО
«Псковская областная типография»,
180004, г. Псков, ул. Ротная, д. 34

ISBN 978-5-299-00858-6

© Кельмансон И. А. (текст), 2017
© ООО «Издательство „СпецЛит“», 2017

ОГЛАВЛЕНИЕ

Условные сокращения	5
Введение	6
Глава 1. Подготовительный этап исследования	14
Глава 2. Дизайн исследования	34
2.1. Описательные исследования	34
2.1.1. Описания клинических случаев и серии наблюдений	38
2.1.2. Исследование популяционных срезов	39
2.1.3. Корреляционные исследования	66
2.2. Аналитические исследования	77
2.2.1. Обсервационные аналитические исследования	78
2.2.1.1. Исследования по принципу «случай—контроль»	78
2.2.1.2. Когортные исследования	88
2.2.2. Интервенционные исследования	100
2.3. Нерандомизированный (квазиэкспериментальный) и рандомизированный дизайны	108
2.3.1. Нерандомизированный (квазиэкспериментальный) дизайн	108
2.3.2. Рандомизированный дизайн	113
2.4. Открытое, слепое, двойное слепое исследование	122
2.5. Исследования в параллельных, связанных группах и перекрестный дизайн	124
2.6. Экспериментальный дизайн с одним испытуемым	127
Глава 3. Процесс измерения	134
3.1. Методы и методика измерения	136
3.2. Основные группы методов	137
3.2.1. Опросники	140
3.2.2. Интервьюирование	147
3.2.3. Наблюдение	152
3.3. Оценка качества измерений	160
3.3.1. Надежность измерения	160
3.2.2. Валидность измерения	179
3.2.3. Альтернативные подходы к оценке качества измерений	193
Глава 4. Анализ данных	198
4.1. Представление и суммирование данных	198
4.2. Оценка влияния случая на результаты исследования	210
4.2.1. Проверка гипотезы	210
4.2.2. Оценочные методы: определение доверительных интервалов	218

4.3. Выраженность эффекта	224
4.4. Численность выборки и статистическая сила исследования ...	238
Глава 5. Оценка валидности исследования	247
5.1. Ассоциации и причинно-следственные связи	248
5.1.1. Угрозы внутренней валидности	256
5.1.1.1. Наличие систематической ошибки	256
5.1.1.2. Влияние искажающих факторов	264
5.1.1.3. Минимизация влияния искажающих факторов ...	268
5.1.1.3.1. Стандартизация	268
5.1.1.3.2. Ограничение	272
5.1.1.3.3. Матчинг	273
5.1.2. Особенности анализа данных с учетом возможного влияния искажающих факторов	276
5.1.2.1. Особенности анализа данных в случаях матчинга	276
5.1.2.2. Особенности стратифицированного анализа	278
5.1.2.3. Основы многофакторного анализа	282
5.2. Принципы мета-анализа	296
Глава 6. Письменное представление сведений об исследо- вании	301
6.1. Обзор литературы	301
6.2. Материал и методы исследования	302
6.3. Результаты	304
6.4. Обсуждение результатов	306
6.4.1. Значение результатов	307
6.4.2. Достоинства и недостатки исследования	308
6.4.3. Научное и практическое значение	309
Заключение	313
Приложения	316
<i>Приложение 1.</i> Таблица случайных чисел	316
<i>Приложение 2.</i> Критические значения t -распределения Стьюдента для двустороннего критерия	320
<i>Приложение 3.</i> Критические значения распределения χ^2	321
Литература	322

УСЛОВНЫЕ СОКРАЩЕНИЯ

- ДИ – доверительный интервал
- СВСМ – синдром внезапной смерти младенцев
- СО – стандартное отклонение
- СОС – стандартная ошибка среднего

ВВЕДЕНИЕ

Уже лучше совсем не помышлять об отыскании каких бы то ни было истин, чем делать это без всякого метода, ибо совершенно несомненно то, что подобные беспорядочные занятия и темные мудрствования помрачают естественный свет и ослепляют ум. Всякий, привыкший таким образом блуждать во мраке, настолько ослабляет остроту своего зрения, что не может больше переносить яркий свет.

Рене Декарт. Правила для руководства ума

От метода, от способа действия зависит вся серьезность исследования. Все дело в хорошем методе. При хорошем методе и не очень талантливый человек может сделать много. А при плохом методе и гениальный человек будет работать впустую и не получит ценных, точных данных. Этой истиной вы должны проникнуться.

И. П. Павлов. Лекции по физиологии

Нет сомнения в том, что как сами исследования, так и их методика в настоящее время не отличаются изяществом, но со временем они усовершенствуются.

*К. Роджерс. О становлении личности.
О психотерапии глазами психотерапевта*

Клиническая психология — наука, изучающая психологические особенности людей, страдающих различными заболеваниями, методы и способы диагностики психических отклонений, дифференциации психологических феноменов и психопатологических симптомов и синдромов, психологию взаимоотношений пациента и медицинского работника, психопрофилактические, психокоррекционные и психотерапевтические способы помощи пациентам, а также теоретические аспекты психосоматических и соматопсихических взаимовлияний (Менделевич В. Д., 2001). Наука может успешно развиваться только в том случае, если она использует для накопления фактов научно обоснованные подходы и клиническая психология как наука реализует в исследованиях общие принципы научного познания. Методология — это система принципов и способов организации и построения теоретической и практической деятельности, объединенная учением

об этой системе (Карвасарский Б. Д., 2011). Методология включает конкретно-научные приемы исследования, которые, в свою очередь, реализуются в специальных процедурах — методиках получения научных данных. Являясь психологической дисциплиной, клиническая психология во многом опирается на методологию и методы общей психологии. В своей научной работе исследователь всегда реализует ту или иную методологию. Для того чтобы последнее происходило последовательно и плодотворно в любой области психологии, весьма существенно, чтобы исследователь осознанно подходил к ее выбору и использованию.

В отличие от других естественных наук, психология имеет дело с таким неоднозначным объектом, как психика вообще и человеческая психика в частности, поэтому психические явления, изучаемые клинической психологией, часто предстают как очень вариативные, не обладающие выраженной повторяемостью и устойчивостью (Карвасарский Б. Д., 2011). В клинической психологии, как и в медицине, выявляемые закономерности чаще всего носят вероятностный (стохастический), а не детерминистский характер. Такая вероятностная модель во многом предопределяет обоснованность использования эпидемиологических принципов в организации клиничко-психологического исследования. На сегодняшний день отсутствует общепринятое определение медицинской эпидемиологии. Например, ее определяют как раздел медицинской науки, задачей которого является изучение распространенности заболеваний и факторов, влияющих на риск их возникновения в популяции (Hennekens С. Н. [et al.], 1987). Эпидемиология базируется на двух важнейших предпосылках:

- заболевания (патологические состояния) не возникают случайно;

- риск возникновения заболевания (патологического состояния) определяется взаимодействием факторов, повышающих и понижающих его вероятность, и идентификация этих факторов возможна в ходе систематических исследований различных популяций или групп индивидуумов, в различных географических условиях и в различные периоды времени.

Такое концептуальное понимание предмета эпидемиологии оказало существенное влияние на ее практику. Возникла необходимость разработать новые методологические подходы к планированию и осуществлению эпидемиологических исследований, основанные на использовании всех доступных средств из арсенала биомедицинской статистики, в результате чего достигнут современный уровень медицинской эпидемиологии. С учетом категорий эпидемиологии исследователь начинает с предположений, касающихся возможного влияния определенных факторов на риск возникновения того или иного

феномена. Основанием для подобных предположений могут быть клинический опыт, накопленные сведения о характере заболевания, данные лабораторных исследований, а иногда исключительно теоретические предпосылки. Результатом является формирование рабочей гипотезы, которая подлежит проверке. Систематический сбор данных и их анализ предполагают поиск ответа на вопросы:

— Существуют ли статистически значимые ассоциации между изучаемыми факторами и риском заболевания?

— Отличается ли вероятность того или иного исхода в присутствии конкретного фактора или воздействия от соответствующей вероятности в условиях отсутствия этого фактора или воздействия?

Следующим этапом является оценка значимости выявленной ассоциации, для чего исключают возможные альтернативные объяснения, например простое удачное стечение обстоятельств (шанс), систематические ошибки, допущенные в ходе сбора или интерпретации данных, а также неучтенные дополнительные факторы, которые, в свою очередь, также могут быть ответственны за выявленную ассоциацию. На последнем этапе выносится решение о том, отражает ли выявленная ассоциация причинно-следственную связь между воздействием и заболеванием, что выходит за рамки однократного исследования и предполагает учет таких показателей, как сила выявленной связи, принципиальное совпадение полученных результатов с данными других исследователей, а также обоснованность предполагаемой причинно-следственной связи с общебиологических позиций. Такая стадийность эпидемиологического мышления во многом совпадает и с исторической эволюцией учения об эпидемиологии.

Базисными науками для клинической психологии являются обшая психология и психиатрия. Именно на стыке этих наук и возникла клиническая психология. Предметом научных исследований клинической психологии и психиатрии являются психические расстройства; клиническая психология, кроме того, занимается такими нарушениями, которые по своей значимости не равноценны болезни, а также психическими аспектами соматических расстройств (Карвасарский Б. Д., 2011). Некоторые исследователи полагают, что клиническая психология и психиатрия близки, даже если учесть разные подходы к предмету исследования. Психиатрия, как частная область медицины, больше учитывает соматический аспект психических расстройств; в клинической психологии основными являются психологические аспекты (Перре М., Бауманн У., 2006). Следовательно, подходы, разрабатываемые в клинической психологии и психиатрии, иногда не имеют выраженных различий и нередко реализуются в совместных исследованиях. Показательно, что эпидемиологические принципы исследований получили широкое распро-

странение в психиатрии (Cooper B., Eastwood M. R, 1992; Mezzich J. E. [et al.], 1994; Susser E. S., 2006; Tsuang M. T. [et al.], 2011).

С другой стороны, клиническая психология оказывает влияние на развитие теории и практики психиатрии, неврологии, нейрохирургии, внутренних болезней, педиатрии и других медицинских дисциплин. Следовательно, методология клиничко-психологического исследования должна быть таковой, чтобы итоги исследования были максимально совместимы с теми методологическими принципами, которые используются и в этих научных дисциплинах.

Принципиальной особенностью клиничко-психологических исследований является их проведение в человеческой популяции. В фундаментальных дисциплинах возможны многочисленные лабораторные исследования и эксперименты на животных, позволяющие в максимальной степени контролировать выраженность и характер изучаемых воздействий, однако в человеческой популяции подобные подходы оказываются возможными лишь в ограниченных ситуациях. Более того, результаты экспериментальных исследований на животных редко могут быть в полной мере экстраполированы на человека. Вот почему нередко лишь эпидемиологические данные позволяют количественно оценить связь между уровнем и характером предполагаемого патогенного или лечебно-профилактического воздействия, с одной стороны, и эффектом, реализующимся в виде увеличения или снижения риска патологических нарушений, — с другой, а также дают возможность модифицировать популяционный риск за счет активного вмешательства. Правильно спланированные и четко реализованные исследования в сочетании с последующим адекватным анализом и интерпретацией полученных результатов чрезвычайно эффективны при выявлении причин заболеваний, психологических нарушений и разработке мер их профилактики.

Предметом профессиональной деятельности клинического психолога являются психические процессы и состояния, индивидуальные и межличностные особенности, социально-психологические феномены, проявляющиеся в различных областях человеческой деятельности. Клинический психолог выполняет диагностическую, экспертную, коррекционную, профилактическую, реабилитационную, консультативную, научно-исследовательскую и некоторые другие виды работ. В ходе повседневной профессиональной деятельности квалифицированный клинический психолог неизбежно сталкивается с необходимостью найти и применить на практике те средства, которые могут быть полезны для конкретного пациента. Подобный поиск предполагает наличие индивидуального клинического опыта и доступность максимально информативных сведений из других источников, подвергшихся систематическому и критичному отбо-

ру и осмыслению. Индивидуальный опыт понимается как степень профессиональной зрелости специалиста и совокупность тех взглядов, которые формируются у него в ходе клинико-психологической практики. Богатый практический опыт главным образом реализуется в виде максимально эффективной диагностики и способности учитывать индивидуальные качества пациента в ходе принятия тех или иных решений. Сведения из других источников представляют собой результаты исследований, которые можно использовать в отношении конкретного пациента, будь то проблемы профилактики, диагностики, лечения, реабилитации. В таком качестве могут быть приняты результаты фундаментальных исследований или — чаще всего — прикладных клинических и клинико-психологических исследований в медицинских центрах, где получают лечение пациенты с соответствующей патологией. Такие прикладные исследования направлены на изучение эффективности методов диагностики, значимости отдельных прогностических признаков, эффективности и безопасности диагностических, лечебных и реабилитационных мероприятий. Данная информация имеет наибольшее значение для практического клинического психолога, так же как и для практического врача.

Существенной проблемой является то, что данные из других источников могут противоречить сложившимся представлениям и стереотипам специалиста, а в ряде случаев даже дискредитировать подходы к диагностике и лечению, которые он привык использовать, и тем не менее предлагать новые схемы, которые могут оказаться более эффективными и безопасными. Грамотный специалист, не утративший способности к критичному мышлению, использует как собственный клинический опыт, так и сведения, полученные со стороны, не ограничиваясь одним источником. С одной стороны, клинический опыт нужен для того, чтобы практическая работа клинического психолога не пострадала от достаточно абстрактной информации из разных источников, так как даже наиболее обоснованные меры могут оказаться неприемлемыми в конкретных ситуациях и у конкретных пациентов. С другой стороны, без должного учета внешних аргументов специалист рискует отстать от жизни, а предложенное им лечение может оказаться не только малоэффективным, но и опасным.

В сложившейся ситуации непрерывное самообразование и анализ доступной информации — неотъемлемые составляющие профессиональной деятельности. Становится актуальным вопрос о качестве той информации, которую специалист может найти в разных источниках. Можно сказать, что в современных условиях специалист страдает не столько от дефицита информации, сколько от невозможности или неумения отобрать и проанализировать ее. Даже в тех случаях, когда специалист имеет возможность познакомиться с интересными дан-

ными, опубликованными в солидных журналах и монографиях, он рискует неверно истолковать эти данные или необоснованно применить их на практике, в контексте тех лечебно-диагностических проблем, которые связаны с конкретным пациентом. С другой стороны, специалист рискует проигнорировать результаты сторонних исследований, которые могут иметь принципиальное значение для решения стоящих перед ним вопросов. У данного парадокса есть несколько причин:

- неумение верно и предельно четко сформулировать вопрос, на который специалист хочет получить ответ;

- появление в клинической практике доказательств нового типа, к правильному восприятию которых клинический психолог может быть не подготовлен.

Прежде всего речь идет о рандомизированных контролируемых испытаниях, получающих все большее распространение во всех сферах медицины, их результаты в ряде случаев способствовали пересмотру традиционных лечебных и диагностических подходов.

До настоящего времени продолжаются дискуссии по поводу того, в какой мере клинический психолог в своей профессиональной деятельности должен ориентироваться исключительно на научные данные. Некоторые авторы вообще высказывают сомнение в том, что клиническая психология может быть достаточно эффективной, если опирается исключительно на науку. Важнейшие аргументы скептиков сводятся к следующему:

- Данные, полученные в больших группах испытуемых, не могут быть использованы при работе с конкретными пациентами.

- У пациента имеются проблемы в конкретный момент, и у клинического психолога нет возможности ждать, когда будет проведено научное исследование, которое даст ответы на возникшие вопросы.

- Каждый индивидуум уникален, так как имеет свой жизненный опыт, привержен конкретной культуре, находится в своем социальном контексте. В связи с этим маловероятно, чтобы общепсихологические принципы были полезны для работы с любым конкретным пациентом.

- В настоящее время отсутствуют научно обоснованные доказательства, позволяющие понять и устранить многие психологические проблемы пациентов, наблюдаемые клиническими психологами (Hunsley J., Lee C. M., 2010).

Несмотря на то что некоторые из приведенных аргументов не лишены оснований, полностью согласиться с ними означало бы признать, что работа клинического психолога сводится исключительно к ориентации на собственные убеждения, интуицию или личный опыт, а сама профессия клинического психолога — это не научная дисциплина.

лина. По-видимому, именно такой взгляд на психологию дал повод известному британскому ученому-химику Энтони Стандену горько пошутить: «Психологи на словах придерживаются научного метода и используют его, когда это удобно, однако когда это неудобно, они дают волю своей необузданной фантазии» (Standen A., 1950). Ричард МакФолл, будучи президентом Американского научного общества клинических психологов, в 1991 г. опубликовал манифест, призывавший клинических психологов в своей практической деятельности ориентироваться исключительно на методы, безопасность и эффективность которых научно подтверждены (McFall R. M., 1991). Эти положения были дополнены и развиты в дальнейшем (Baker T. V. [et al.], 2008). В соответствии с высказанными положениями, научная клиническая психология является единственно приемлемой формой. Данное положение позволяет сделать ряд выводов.

Во-первых, за исключением тех случаев, когда изучается эффективность психологической помощи в ходе строго регламентированных экспериментальных исследований, допустимо предложить психологические услуги населению только после того, как они удовлетворяют следующим четырем минимально необходимым критериям:

- должна быть четко описана суть оказываемой психологической помощи;
- должна быть недвусмысленно названа предполагаемая польза от оказания психологической помощи;
- заявленная польза от оказания психологической помощи должна быть научно обоснованной;
- возможные негативные побочные эффекты, которые могут перевесить предполагаемую пользу, должны быть исключены на основании эмпирических данных.

Во-вторых, первоочередной и важнейшей целью программ подготовки клинических психологов должна быть ориентация на формирование их максимальной научной компетентности.

На сегодняшний день сложилось специальное направление медицины — доказательная медицина (evidence-based medicine), задача которой — обоснование методологических и практических подходов, призванных помочь специалисту в поиске и критичном осмыслении медицинской информации. **Доказательная медицина** — система научно обоснованных подходов, направленных на оптимизацию диагностических и лечебно-профилактических мероприятий в отношении конкретных пациентов, она требует четкой постановки вопросов, использования максимально доказательной доступной медицинской информации и ее критичного осмысления (Власов В. В., 2001; Кельмансон И. А., 2004; Straus S. E., 2011). Принципы доказательной медицины получили широкое применение в различных

медицинских специальностях, в том числе в психиатрии (Gray G. E., 2004; Taylor C. B., 2010; Pompili M., Tatarelli R., 2011). Эти принципы все шире внедряются в теорию и практику клинической психологии (L'Abate L., 2008; Hunsley J., Lee C. M., 2010). Можно говорить о формировании доказательной клинической психологии, и знание ее принципов является обязательным не только для практикующих клинических психологов, но и для ученых, генерирующих эту информацию в ходе исследований. Когда ученый планирует исследование, определяет контингент обследуемых и критерии их отбора, методику анализа изучаемого эффекта и иные компоненты работы, которые в современной литературе принято называть дизайном исследования, ему нужно отдавать себе отчет в том, что ошибки и просчеты, допущенные им на этом этапе, неизбежно приведут к недоказательным, необоснованным выводам, которые могут оказаться не только бесполезными, но и вредными.

С сожалением приходится констатировать неудовлетворительную осведомленность многих клинических психологов — как исследователей, так и практиков — об основополагающих методологических подходах, используемых в современной клинической психологии. Данное обстоятельство приводит к необоснованному сужению круга исследовательских возможностей, которые могли быть успешно реализованы. Сказанное диктует настоятельную необходимость ознакомить клинических психологов с методологией исследования в клинической психологии, основу которой во многом составляют современные эпидемиологические подходы и принципы доказательности. Руководствуясь высказанными соображениями, автор взял на себя смелость предложить данную книгу широкому кругу клинических психологов. Предполагается, что читатель знаком с основами статистики, которые пригодятся для понимания принципов методологии исследования. В книге приведен ряд статистических и эпидемиологических терминов, математических формул, использование которых неизбежно в рамках обсуждаемой темы, тем не менее автор стремился изложить материал максимально доступно, ориентируясь на массовую аудиторию клиницистов и клинических психологов.

Автор рассчитывает на заинтересованное отношение читателей к представленной на их суд книге, с пониманием и благодарностью примет все конструктивные замечания и предложения.

ПОДГОТОВИТЕЛЬНЫЙ ЭТАП ИССЛЕДОВАНИЯ

Научные исследования, в том числе в области клинической психологии, проводятся прежде всего в интересах практики и дальнейшего развития теории. Они предпринимаются также для преодоления определенных трудностей в процессе познания новых явлений, объяснения ранее неизвестных фактов или для выявления неполноты старых способов объяснения известных фактов. Научную проблему часто и справедливо характеризуют как осознанное незнание. Действительно, пока мы не осознаем, что не знаем о каких-либо явлениях или об их сторонах, нам все ясно. Дальнейшее движение по пути познания возможно только после признания, что в наших знаниях есть проблемы, заполнить которые удастся в результате дальнейшего развития науки и успешных действий на практике. Итак, **научная проблема** — это форма научного знания, содержание которой составляет то, что еще не познано человеком, но что нужно познать, т. е. это знание о незнании, это вопрос, возникший в ходе познания или практики и требующий научно-практического решения. У проблемы есть эмпирический и теоретический аспекты (Баскаков А. Я., Туленков Н. В., 2004).

Постановка проблемы и определение тематики исследования

Постановка проблемы означает:

- определение известного и неизвестного, объясненных фактов и фактов, требующих объяснения; фактов, соответствующих теории и противоречащих ей;
- формулировку вопросов, выражающих основной смысл проблемы, обоснование их правильности и важности для науки и практики;
- определение конкретных задач исследования, последовательности их решения и **методов**, которые будут применяться при этом (Баскаков А. Я., Туленков Н. В., 2004).

Прежде всего необходимо провести предварительный анализ, в процессе которого будет сформулирована сама проблема, намечены пути ее решения и определена конкретная тема исследования. Существенная роль отводится изучению научной литературы, которое идет параллельно с формулировкой вопросов исследования. Анализ литературы позволяет:

— оценить, насколько хорошо и полно исследована интересующая автора тема, какие ее аспекты требуют изучения, достаточно ли полно описан интересующий исследователя феномен;

— оценить, насколько полно в литературе представлены возможные ответы на вопросы, которые исследователь ставит перед собой:

Что может настоящее исследование добавить к пониманию проблемы?

Требуются ли дополнительные исследования по теме?

Проводились ли ранее аналогичные исследования?¹

— помочь сформулировать вопросы исследования в свете имеющихся научных теорий или предыдущих работ и, возможно, сформулировать ту теоретическую и концептуальную основу, на которой будет строиться исследование;

— помочь с выбором методов исследования и планированием его дизайна;

— оценить, какие параметры исследовались в ходе предыдущих работ, какие использовались подходы, в чем достоинства и недостатки дизайнов предыдущих исследований (Barker C. [et al.], 2002).

Знакомство с литературой не следует рассматривать как непосредственную задачу исследования: оно является непреложным условием подготовки и планирования работы. Здесь возможны два исключения:

— анализ литературы является самостоятельным разделом исследования (например, изучение массива публикаций по той или иной теме за определенный период времени);

— проводится мета-анализ, когда каждая публикация по теме является непосредственно объектом научной разработки.

Изучение литературы по проблеме исследования стоит начать со знакомства с электронными библиографическими базами данных, которые в последние годы привлекают к себе особое внимание как к средству эффективного библиографического поиска. Существующие электронные библиографические базы данных, как правило, построены по сходным принципам. Каждый литературный источник обычно содержит три основных элемента: собственно библиографические сведения (название работы, имена и фамилии авторов, название книги или журнала, год издания, том, номер, страницы и т. п.), резюме данной публикации (Abstract), а также ключевые слова, или дескрипторы, присвоенные данной публикации. Ключевые слова выбраны из тех, которые приводятся в тезаурусе (словаре).

¹ Необходимо помнить, что даже при наличии аналогичных исследований вероятное дублирование никак не дискредитирует идею исследования, поскольку маловероятно, чтобы два независимых исследователя организовали свои исследования по абсолютно идентичному дизайну и на абсолютно идентичном материале.

Существует несколько способов получить доступ к подобным электронным базам данных.

Традиционный доступ к бумажным носителям информации. В качестве примеров можно назвать:

— библиографические обзоры в еженедельном сборнике Current Contents, в котором Институт научной информации (США) приводит сведения о текущих публикациях по различным разделам науки, в частности в области биомедицинских наук и здравоохранения;

— библиографические обзоры Index Medicus, Excerpta Medica, а также «Медицинский реферативный журнал».

Прямой доступ к электронным базам данных. Его распространение обусловлено широким внедрением электронных технологий, в частности электронных носителей информации и Интернета. Возможен поиск по ключевым словам и по образцу свободного текста. Каждый вариант поиска имеет свои достоинства и недостатки.

Поиск по ключевым словам основан на использовании тезауруса, в котором представлен перечень возможных терминов. *Преимущество* обеспечено стандартизацией важнейших понятий, описанием проблемы в рамках определенных терминов. Кроме того, ключевые слова группируются в определенную иерархическую структуру: на основе одного более емкого, обобщающего термина можно осуществить поиск во всех подразделах, относящихся к этому термину. Возможность реализации подобного поиска предполагает, что каждая статья должна сопровождаться перечнем тех ключевых слов, которые наилучшим образом описывают содержание работы. Это довольно сложная задача, и порой не представляется возможным полностью охарактеризовать содержание работы при помощи нескольких стандартных терминов (в этом можно усмотреть *недостаток*). Нередко требуется определенный срок, прежде чем новые медицинские концепции получат свое отражение в виде определенных терминов и понятий, включенных в тезаурус.

Поиск по свободному тексту предполагает обнаружение любых слов или их комбинаций, отдельных фраз, содержащихся как в названии работы, так и в самом тексте работы. *Преимущество* состоит в том, что гибкий поиск не ограничен рамками тех стандартных терминов, которые заданы тезаурусом. Результаты поиска не зависят исключительно от тех ключевых слов, которые были приведены авторами в статье или резюме. Поиск на основе свободного текста полезен, если по каким-то причинам ключевые слова не охватывают все те аспекты, которые были предметом исследования. *Недостатком* является невозможность использовать синонимы. Если поиск статей основан на обнаружении заданных слов в тексте, то отбор публикаций будет ограничен исключительно этими словами, и публикации,

посвященные той же проблеме, но использующие синонимы тех же понятий, могут быть пропущены. Для того чтобы избежать подобной ситуации, следует проводить поиск, применяя максимально возможную комбинацию близких понятий. В определенных ситуациях весьма затруднительно составить исчерпывающий перечень синонимов, отражающих содержание одной и той же темы или концепции. С другой стороны, подобная работа позволяет исследователю всесторонне освоить все ту терминологию, которая отражает суть изучаемой проблемы, причем добиться этого еще до того, как будут внесены соответствующие коррективы и дополнения в тезаурус библиографической базы данных.

Логические операторы. Редко бывает так, что один заданный термин позволяет провести всеобъемлющий поиск библиографических материалов в рамках заданной темы. Как правило, любое исследование сочетает в себе несколько концептуальных элементов.

Пример

Исследование риска формирования синдрома дефицита внимания и гиперактивности в школьном возрасте у детей, родившихся недоношенными, будет, как минимум, удовлетворять следующим концептуальным критериям поиска: синдром дефицита внимания и гиперактивности, школьный возраст, недоношенные новорожденные. Для того чтобы обеспечить подобный отбор, необходимо использовать логические операторы. Чаще всего применяют логические операторы «И», «ИЛИ», «НЕ». Если задаются условия поиска: А «И» Б, это означает, что отобранные статьи должны одновременно иметь признаки А и Б. Например, если задать в качестве условий поиска «недоношенные новорожденные» «И» «синдром дефицита внимания», то должны быть отобраны публикации, в которых рассматриваются вопросы, связанные с недоношенными новорожденными и синдромом дефицита внимания. При использовании в ходе поиска логического оператора «ИЛИ» будут отбираться работы, в которых содержится либо термин А, либо Б (как отдельно, так и совместно). Например, если задать параметры поиска «недоношенные новорожденные» ИЛИ «синдром дефицита внимания», будут отобраны работы, посвященные недоношенным детям или проблеме синдрома дефицита внимания.

Medline. Одной из наиболее популярных международных электронных библиографических баз данных в сфере биомедицинской литературы является Medline (MEDlars on LINE — доступ он-лайн к MEDLARS (MEDical Literature Analysis and Retrieval System)). В свою очередь, MEDLARS представляет собой электронную библиографическую базу данных, которая по сути и по содержанию эквивалентна издаваемым библиографическим справочникам Index Medicus. Medline формируется Национальной медицинской библиотекой США

(National Library of Medicine, NLM) ¹. Medline обеспечивает доступ ко всем литературным источникам, реферлируемым в Index Medicus, International Nursing Index и Index to Dental Literature. Предметом реферирования являются статьи, публикуемые приблизительно в 4000 периодических медико-биологических изданиях. Примерно в 80 % случаев литературные источники сопровождаются резюме на английском языке. База данных Medline пополняется еженедельно, в результате каждый месяц база увеличивается на 35—40 тысяч работ. Существует несколько способов получить доступ к Medline, например на сайте Национальной медицинской библиотеки (<http://nlm.nih.gov.PubMed>). Информация о публикации, содержащейся в Medline, распределена по отдельным рубрикам (полям). Перечислим их:

- индивидуальный идентификационный номер данной публикации;
- название статьи на английском языке. Если статья была написана не по-английски, в квадратных скобках приводится перевод названия статьи на английский язык, а затем при помощи транслитерации приводится название статьи в оригинале;
- список авторов — фамилии и инициалы;
- адрес для переписки и электронный адрес (при наличии) основного автора статьи;
- собственно библиографическая информация: название журнала (аббревиатура), в котором опубликована статья, место издания, дата его выхода в печать, год, том и номер, интервал страниц;
- международный стандартный номер данного периодического издания (ISSN).

Поле «Примечание» (Note) позволяет указать выявленные ошибки, которые были допущены при реферировании статьи. Также в примечании указывается язык, на котором была написана работа, и язык, на котором написано резюме в оригинале.

Поле «Тип документа» (Document type), или «Тип публикации» (Publication Type, PT), содержит важную информацию с точки зрения литературного поиска. Насчитывается около 40 различных типов публикаций, например клиническое испытание (clinical trial), контролируемое клиническое испытание (controlled clinical trial), обзор (review). Указывается, имеется ли резюме на английском языке для статей, написанных на других языках.

Поле «Анонсирование журнала» (Journal announcement) содержит сведения о дате (годе и месяце), когда сведения об опубликованном журнале поступили в NLM.

¹ Следует отметить, что Национальная медицинская библиотека США (NLM), помимо обсуждавшейся выше Медлайн, поддерживает ряд других библиографических баз данных. Подробное описание этих библиографических баз данных можно найти на сайте NLM (<http://www.nlm.org>).

Поле «Субфайл» (Subfile) содержит информацию о том, к какому реферативному источнику относится данная статья (Index Medicus и т. п.).

Поле «Резюме» (Abstract) содержит резюме статьи и является очень важным с практической точки зрения. Это резюме приводится в том виде, в котором было опубликовано в самой статье. Следует отметить, что Medline допускает возможность хранения в памяти резюме, размер которого не превышает 250 слов, а более длинные резюме приводятся в сокращенном (truncated) виде, при этом указывается: "truncated at 250 words". Вот почему медицинские журналы, реферируемые в Medline, просят авторов при оформлении статей давать резюме не более 250 слов.

Поле «Выбрать тэги» (Check tags) содержит сведения о важнейших характеристических признаках работы.

Поле «Дескрипторы» (Descriptors) предлагает ключевые слова. Речь идет о тех терминах, содержащихся в тезаурусе Medline (Medical Subject Headings, MeSH), которые присваиваются данной конкретной работе индексатором Национальной медицинской библиотеки. В тезаурусе насчитывается примерно 17 000 дескрипторов, расположенных по иерархическому принципу (Lowe H. J., Barnett G. O., 1994). Более общие дескрипторы (например, сердечно-сосудистые заболевания) включают в себя менее общие, более узкие и конкретные понятия (ишемическая болезнь сердца, нарушения ритма сердца и т. п.). Эта структура выстраивается также за счет присвоения терминам, содержащимся в MeSH, соответствующих номеров. Достоинством тезауруса является возможность автоматически осуществить поиск других терминов, каким-то образом связанных с одним относительно известным и достаточно точным термином. Имеются также и тематические подразделы (Topical SubHeading), которые дополняют и уточняют имеющиеся дескрипторы. Например, термин, обозначающий определенное заболевание, может быть дополнен таким образом, что начинает обозначать какой-то отдельный аспект, связанный с этим заболеванием (лечение заболевания, диагностику и т. п.). Важнейшей проблемой, возникающей при работе с MeSH, является необходимость найти именно те термины, которые будут наилучшим образом отражать суть изучаемой проблемы. Для облегчения такого поиска в тезаурус включены синонимы. Эти термины сами по себе не являются непосредственной частью иерархической структуры тезауруса, но помогают адресовать исследователя к нужному термину в MeSH.

PsycINFO (www.apa.org/psycinfo) Американской психологической ассоциации (American Psychological Association, APA) предоставляет разнообразные поисковые средства. Ресурс содержит ссылки на журнальные статьи, диссертации, технические отчеты, книги и главы

из книг. База охватывает литературу более чем 45 стран и включает ссылки на более чем 1300 журналов. Каждый месяц в базу данных добавляется около 5500 новых наименований. База данных построена по тому же принципу, что и Medline: каждая запись содержит несколько важных разделов — полей: название статьи, ее автора (авторов), все необходимые выходные данные (название журнала, номер, страницы), краткий обзор статьи (abstract) и дескрипторы. Поиск статей возможен по определенным терминам, и если исследователь не уверен в том, с каких терминов начать поиск, можно воспользоваться собственным тезаурусом PsycINFO. Тезаурус в данном случае представляет собой электронную версию «Словаря психологических терминов» (Thesaurus of Psychological Index Terms), изданного APA. В нем предложена расширенная система перекрестных ссылок на термины, использующиеся как ключевые слова при поиске.

Имеются и другие библиографические базы данных, связанные с вопросами психологии, в том числе медицинской (EPIC, Psycbase, PsycSCAN).

EMBASE содержит общемедицинскую информацию, является электронной версией реферативных обзоров Excerpta Medica (www.elsevier.com/solutions/embase-biomedical-research). Она содержит библиографические источники — более 3500 наименований журналов, издаваемых в 110 странах. Преимущества базы составляют достаточная полнота охвата европейских периодических медицинских изданий, оперативность реферирования работы в базе. Тезаурус, используемый в этой базе данных, содержит около 39 000 терминов, относящихся к медицинским понятиям и названиям лекарственных препаратов.

BIOSIS содержит общемедицинскую информацию о публикациях в области медицины, биологии, экологии, биохимии, фармакологии. Можно найти информацию о статьях в журналах, монографиях, а также о материалах конгрессов. Каждый год библиография пополняется примерно на полмиллиона литературных источников. BIOSIS используется для подготовки многих реферативных сборников:

— Biological Abstract содержит информацию о статьях, опубликованных в 6500 периодических изданиях, сопровождается и авторскими резюме этих статей;

— Biological Abstract/RRM (Reports, Reviews, Meetings) содержит информацию о сообщениях, представленных на конгрессах.

Доступ к указанной базе данных в настоящее время возможен через Интернет (<http://www.biosis.org>).

PASCAL поддерживается французским Национальным институтом научной и технической информации. Эта библиографическая база данных содержит общемедицинские сведения о мировой литературе в области науки, технологии и медицины, издаваемой на раз-

личных языках. Медицинская литература составляет примерно треть имеющихся публикаций. Еще около четверти публикаций посвящено биологии. Важно отметить, что здесь много работ на русском языке и так называемых «серых» публикаций, т. е. таких работ, которые были опубликованы без соблюдения всех принятых требований, предъявляемым к рецензируемым работам.

LILACS (Literature Latino-Americana y del carib en Ciencias de la Salud) преимущественно охватывает медицинские публикации в странах Латинской Америки.

MEDIDOC — база данных, охватывающая медицинские публикации на французском языке.

SCISEARCH (Science Citation Index) предлагает не только перечень самих публикаций, но и список тех статей, в которых эти публикации цитируются.


Current Contents предназначено для оперативного информирования об опубликованных работах. Издание ориентировано на ознакомление с текущими публикациями (как явствует из самого названия) и в меньшей степени рассчитано на ретроспективный поиск.

SIGLE (System for Information on Grey Literature in Europe) специально предназначена для обеспечения доступа к «серой» литературе, опубликованной в Европе. В указанную категорию попадают научные, технические и экономические отчеты, подготовленные в общественном или частном секторе; диссертации, сообщения на конференциях, рекомендации и методические указания, некоторые официальные документы, техническая документация. SIGLE создана в 1980 году и охватывает академические и прикладные научно-технические исследования, а также работы в области гуманитарных и социальных наук.

E-library (<http://elibrary.ru>) — российская научная электронная библиотека и связанный с ней Российский индекс научного цитирования (РИНЦ) заслуживают особого интереса. РИНЦ — это национальная информационно-аналитическая система, аккумулирующая более 7 миллионов публикаций российских авторов, а также информацию о цитировании этих публикаций более чем в 4500 российских журналах. Она предназначена для оперативного обеспечения научных исследований актуальной справочно-библиографической информацией, является мощным инструментом, позволяющим осуществлять оценку результативности и эффективности деятельности научно-исследовательских организаций, ученых, уровень научных журналов и т. д. Основу системы составляет библиографическая реферативная база данных, в которой индексируются статьи в российских научных журналах. В последние годы в РИНЦ стали включаться также и научные публикации других типов: доклады на конференциях, монографии, учебные пособия, патенты, диссертации. База содержит

следующую информацию: выходные данные, авторы публикаций, места их работы, ключевые слова и предметные области, а также аннотации и пристатейные списки литературы. Хронологические рамки системы — с 2005 года по настоящий день, по многим источникам глубина архивов больше. Каждый день в РИНЦ добавляется более 3000 новых описаний публикаций российских ученых. Более 3600 из 4500 российских журналов, обрабатываемых в РИНЦ, представлены в полнотекстовом виде на платформе eLIBRARY.RU, в том числе 3000 журналов — в открытом доступе, что позволяет ознакомиться и с текстом оцениваемой публикации.

Google Scholar (<http://scholar.google.ru>) — интернет-ресурс, который позволяет найти полнотекстовые статьи, а не только рефераты. Поиск осуществляется в приоритетном порядке: первыми отбираются те работы, которые максимально соответствуют заданным параметрам поиска и чаще всего цитируются. В связи с последним обстоятельством Google Scholar полезен при поиске ключевых или классических работ в рассматриваемой области. Сама процедура поиска в Google Scholar довольно проста: она напоминает поиск в Google. Достаточно указать ключевые слова поиска и затем просмотреть результаты этого поиска. Следует отметить, что часто отбираются работы, не вполне соответствующие тематике поиска, поэтому результат отбора стоит дополнительно просмотреть. В связи с тем, что Google Scholar анализирует институтские репозитории, образовательные сайты, сайты издательств и другие доступные базы данных, в процессе поиска возможно дублирование одних и тех же работ, хотя в ряде случаев поисковая система способна «объединять» такие дубликаты в единую ссылку. Достоинством является поиск литературы, опубликованной на различных языках, в том числе на русском. В ряде случаев приводятся гиперссылки на полнотекстовую публикацию, однако трудно найти гиперссылки на старые журнальные статьи, монографии, диссертации. В то же время поисковая система позволяет выявить ближайшие библиотеки, в которых можно найти интересующую книгу (Library Search). Сама процедура поиска довольно проста: достаточно указать интересующие исследователя ключевые слова, не обязательно вводить логический оператор «AND» между этими словами, так как Google Scholar автоматически выполняет поиск по всем указанным ключевым словам. Использование при поиске кавычек позволяет искать публикации по заданной фразе и сузить поиск. Поисковая система также позволяет ограничить поиск с учетом даты, автора, названия журнала. Имеется и возможность загружать выявленные работы в библиографические приложения для компьютера (EndNote, Reference Manager, RefWorks). Ниже приведен фрагмент поиска публикаций по теме «расстройства сна у детей» начиная с 2011 г. (рис. 1).




Scholar

"расстройство сна у детей"

9 results (0.04 sec)

My Citations



Tip: Search for English results only. You can specify your search language in Scholar Settings.

Articles

Case law

My library

Any time

Since 2015

Since 2014

Since 2011

Custom range...

Sort by relevance

Sort by date

include patents

include citations

Create alert

Расстройство сна у детей с умственной отсталостью
 И.А. Кельмансон - Вопросы современной педиатрии, 2014. - №3. spr-journal.ru
 Аннотация Рассматривается ассоциация расстройства сна с умственной отсталостью у детей. Обращается внимание на непосредственную связь нейрофизиологии сна и формирования интеллекта ребенка, расстройства сна с патологией умственной ...
 Cited by 1. [Cite](#) [Save](#)

РИТМ Оценка качества и структуры сна у детей с патологией ЛОР-органов, бронхиальной астмой и наличием распространенными болезнями нервной ...
 Э.К.А. АБШИДЗЕ... - Педиатрическая ..., 2015. - субрегионика.ru
 ... диагностика расстройства сна и синдрома лунга сна в амбулаторной педиатрической ...
 в организации [Б]. В то же время **расстройства сна у детей и подростков** с ...
[Related articles](#) [All 2 versions](#) [Cite](#) [Save](#) [More](#)

ЭКОЛОГИЧЕСКИЕ И КЛИНИКО-БИОЛОГИЧЕСКИЕ АСПЕКТЫ НАРУШЕНИЙ ЦИРКАДΙΑННЫХ РИТМОВ СОН-БОДРОСТОВАНИЕ У ДЕТЕЙ И ПОДРОСТКОВ
 И.А. Кельмансон - Бюллетень, 2015. - 21bs su
 ... Ключевые слова: мелатонин сон фотопериод хронотип циркадные ритмы. Полный текст: PDF Литература. Кельмансон И. А. **Расстройство сна у детей с умственной отсталостью** // Вопросы современной педиатрии. - 2014. - Т. 13, № 6. - С. 15-20. ...
 Cited by 2. [Cite](#) [Save](#) [More](#)

Расстройство сна у детей с умственной отсталостью (И.А. Кельмансон)
 И.А. Кельмансон - Вопросы современной педиатрии, 2014. - spr-journal.ru
 Аннотация Рассматривается ассоциация расстройства сна с умственной отсталостью у детей. Обращается внимание на непосредственную связь нейрофизиологии сна и

PDF

PDF

from suberegionika.ru

HTML

from suberegionika.ru

Рис. 1. Фрагмент литературного поиска в электронной базе данных Google Scholar

Справа от некоторых публикаций приведены гиперссылки на доступные полнотекстовые версии выявленных статей.

Большая часть научных статей построена по единому принципу: они содержат краткое изложение (резюме) работы, введение, описание материалов и методов исследования, результаты исследования и их обсуждение.

Д. Гудвин (2004) дает рекомендации, направленные на обеспечение максимальной эффективности чтения литературы:

- Прочитайте краткое изложение несколько раз и как можно внимательнее. Оно представляет собой общее описание исследования и, возможно, является самым простым для чтения и понимания.

- Найдите общую формулировку исследуемой проблемы в одном или двух первых абзацах.

- Найдите развернутую формулировку гипотезы или гипотез, проверяемых в исследовании, в конце введения, скорее всего в последнем абзаце. Эти гипотезы основываются на формулировке проблемы и исследовательских вопросах, поставленных в работе и описанных в средней части введения. Запишите гипотезы и всегда помните о них во время чтения.

- Уделите особое внимание описанию процедуры исследования и экспериментальному плану в разделе «Методы». Представьте себя на месте участников и сформируйте четкие представления о том, что они должны делать в ходе исследования.

- Приложите максимум усилий при освоении раздела «Результаты», он может оказаться особенно трудным, так как он содержит статистическую информацию и обозначения, которые могут оказаться незнакомыми. Хорошо написанный раздел «Результаты» должен содержать ясное словесное описание полученных результатов, а также графики и таблицы, облегчающие понимание. Он должен дать возможность понять суть происходящего в ходе исследования без опоры на числа и условные обозначения.

- Раздел «Обсуждение» часто начинается с общего описания результатов, так что если раздел «Результаты» остался непонятен, еще не все потеряно. Главная задача этого раздела — объяснить результаты с точки зрения исходных гипотез, поэтому автор будет выстраивать параллели с разделом «Введение». Стоит обратить внимание на описание того исследования, которое следовало бы провести впоследствии (вопрос «Что дальше?»). Раздел «Обсуждение» является прекрасным источником идей для новых исследований.

Ограничение исследования рамками поставленных вопросов или гипотезы

Данный раздел подготовительного этапа принципиально важен, поскольку методы исследования естественным образом обусловлены сформулированными вопросами. Как отмечалось, при чтении публикаций в

первую очередь также следует обращать внимание именно на вопросы, ответы на которые было призвано дать соответствующее исследование. В хорошо написанных статьях вопросы или гипотеза исследования четко сформулированы, и они, как правило, приводятся в конце раздела «Введение». Бывает и так, что читателю приходится самостоятельно определять вопросы исследования по ходу чтения работы.

Для эффективной работы с библиографией требуются специальные средства:

- традиционный ручной поиск (например, при помощи библиографических карточек);

- специальные прикладные компьютерные программы, предназначенные для работы с библиографической информацией.

В качестве примера последних можно назвать коммерческую программу EndNote. Существует и доступный некоммерческий аналог — программа Mendeley. Mendeley доступен для индивидуальных пользователей в виде приложения для PC и Mac, в виде онлайн-приложения для всех современных браузеров (www.mendeley.com). С помощью этой библиографической программы можно читать и аннотировать статьи в своей библиотеке на домашнем компьютере, через сайт или мобильное приложение. Можно осуществлять работу над статьями совместно с коллегами в открытых и закрытых группах. Принципиально важно, что программа позволяет формировать ссылки на литературные источники по ходу написания текста работы и список литературы, используя один из 6750 встроенных в Mendeley стилей (в том числе доступен стиль цитирования, соответствующий требованиям ГОСТа); любой стиль можно отредактировать или создать свой собственный стиль оформления ссылок.

Весьма значимым является описание степени разработанности выбранной темы, для чего на основе изученной литературы составляется литературный обзор, который должен показать, что именно данная тема еще не раскрыта (раскрыта лишь частично или в другом аспекте) и поэтому нуждается в дальнейшей разработке. Обзор литературы по теме должен показать основательное знакомство исследователя со специальной литературой, умение систематизировать источники, критически их рассматривать, выделять существенное, оценивать ранее сделанное другими исследователями, определять главное в современном состоянии изученности темы.

Актуальность — обязательное требование к исследованию. Поэтому исследователю нужно обосновать актуальность выбранной темы.

Объект — это процесс или явление, порождающие проблемную ситуацию и избранные для изучения.

Предмет — это то, что находится в границах объекта. Объект и предмет исследования как категории научного процесса соотносятся

между собой как общее и частное. В объекте выделяется та его часть, которая служит предметом исследования. Именно на него и направлено основное внимание исследователя, и предмет исследования определяет тему работы. Формулировка объекта и предмета исследования является обязательным элементом планируемого исследования.

Цель исследования — то, что в самом общем виде нужно достигнуть в итоге всей работы. Она устанавливается с учетом объекта и предмета исследования, сформулированной проблемы и доказательства того, что та часть проблемы, которая является темой исследования, еще не получила своей разработки и освещения в специальной литературе. Как отмечают некоторые специалисты, формулировка этого важнейшего элемента работы не может начинаться словами «изучение», «исследование», «анализ» и т. п. Изучение чего-либо не может быть целью работы, целью является установление каких-то закономерностей, действующих в окружающем мире, а изучение, исследование — только средство для этого. Более применимы такие выражения, как «раскрыть» (механизм), «установить» (характер, зависимость), «получить», «создать», «обосновать» (например, эффективность применения того-то), «разработать», «определить» (например, информативные критерии для оценки того-то) и т. п. Можно выбрать иную грамматическую форму, тогда «определить» превратится в «определение», «разработать» — в «разработку» и т. д. (Марьянович А. Т., Князькин И. В., 2009).

В рамках поставленной цели исследователю уже в самом начале следует попытаться хотя бы предварительно сформулировать несколько важнейших вопросов о том, что исследователь хотел бы выявить в ходе планируемой работы. Нужно спросить себя: что является главным ожидаемым итогом исследования? Количество и сложность сформулированных вопросов будут зависеть от масштаба предполагаемого исследования, времени его проведения и доступных ресурсов. Необходимо помнить, что исследование должно чему-то научить самого исследователя: бессмысленно проводить работу, которая лишь подтвердит то, что автор уже знал изначально. Любое исследование подвергает испытанию ожидания автора: они могут как оправдаться, так и быть опровергнуты. Вот почему всегда разумно заранее «проиграть сценарии» исследования, т. е. заранее продумать, каковы возможные результаты и какие уроки исследователь сможет извлечь из них, как в зависимости от результатов изменится представление исследователя о теме (Patton M. Q., 2012). Хорошо организованные исследования дают ценную информацию независимо от того, оправдывают ли их результаты исходные ожидания. Обычно на подготовительной стадии бывает трудно четко сформулировать вопросы, стоящие перед исследованием: достаточно хотя бы обозначить

их в общем виде и в дальнейшем конкретизировать, в этом помогут ознакомление с литературой, консультирование с коллегами, уточнение имеющихся технических возможностей. Как правило, уточнение вопросов занимает по времени несколько месяцев.

Можно выделить несколько типов формулируемых вопросов (Barker C. [et al.], 2002):

— *Вопросы-описания (дескриптивные)*:

Что представляет собой феномен X?

Каковы его признаки, характеристики, вариации?

Как часто он наблюдается?

Примерами могут служить следующие вопросы:

Какую пользу видят пациенты, страдающие биполярным расстройством, от участия в группах взаимной поддержки?

Насколько безопасно чувствуют себя пациенты в психиатрических стационарах?

Как часто в популяции наблюдаются симптомы депрессивных расстройств?

Описательные вопросы обычно направлены на формирование целостной картины того или иного феномена, состояния. Они могут быть сфокусированы на его происхождении и развитии, выявлении типичных примеров. Описательные вопросы могут касаться распространенности того или иного состояния.

— *Вопросы-описания-сопоставления (дескриптивно-сопоставительные)*:

Отличается ли группа X от группы Y? Например, отличаются ли мужчины и женщины по способностям выражать эмоции?

Чем отличаются по внутрисемейным отношениям семьи, где воспитываются мальчики с агрессивным поведением, от семей, где воспитываются мальчики с неагрессивным поведением?

Различается ли частота случаев сексуального насилия в детском возрасте у лиц с симптомами булимии или без них?

Подобные вопросы являются продолжением обычных описательных. Следует отметить, что они не затрагивают проблемы причинно-следственных связей, хотя последние могут предполагаться, как в примере с булимией. Данные вопросы предполагают сопоставление двух групп или более, которые формируются на основе заранее определенных различий по полу, социально-экономическому статусу, диагностической категории. Однако такие вопросы не предполагают учет возможных различий, возникающих в ходе интервенционных вмешательств (например, лечения), соответствующие вопросы относятся уже к категории причинных (каузальных).

— *Корреляционные вопросы:*

Отмечается ли связь между X и Y ?

Влияет ли на эту связь третья переменная Z ?

Примерами могут служить следующие вопросы:

Связан ли уровень супружеской поддержки со сроками выздоровления от депрессивных расстройств?

Имеется ли связь между профессиональным выгоранием у медсестер, работающих в психиатрических стационарах, и случаями физического нападения на них?

Влияет ли этническая принадлежность школьников на связь школьной успеваемости и самооценки?

Корреляционные вопросы фокусируют внимание на возможной ассоциации между двумя переменными и более. Такие ассоциации часто прогнозируемы, исходя из теоретических данных или предыдущих исследований. Хотя корреляционные вопросы затрагивают проблему каузальности и обусловлены интересом к наличию такой возможности, недопустимо использовать их для выявления причинно-следственной связи.

— *Каузальные вопросы:*

Приводит ли X к изменению Y ?

Приводит ли воздействие X к большему изменению Y по сравнению с воздействием Z ?

Примеры подобных вопросов:

Повышает ли тренинг родителей качество воспитания?

Снижает ли семейная терапия частоту обострений заболевания у лиц с диагностированной шизофренией?

Приводит ли использование ноотропных препаратов к улучшению показателей памяти?

Каузальные вопросы шире дескриптивно-описательных и корреляционных, поскольку они направлены на выявление причинно-следственных связей. Выявление таких причинно-следственных связей требует экспериментального исследования в той или иной форме (Cook T. D., Campbell D. T., 1979; Shadish W. R. [et al.], 2001). Некоторые каузальные вопросы формулируются так, что они явно (эксплицитно) предполагают определенное сопоставление, например сопоставление эффективности различных видов терапии у пациентов с определенной патологией. В ряде случаев такое предполагаемое сопоставление носит неявный (имплицитный) характер. В более сложных ситуациях вопросы затрагивают взаимодействие различных переменных (Meltzoff J., 1998; Meltzoff J., Kornreich M., 2008). Например: является ли когнитивная терапия более эффективной по сравнению с традиционной терапией пограничных личностных расстройств у мужчин, но не у женщин?

— *Вопросы, касающиеся измерений:*

Насколько удовлетворительно (надежно, валидно и эффективно) может феномен X быть измерен при помощи измерения M ?

Насколько удовлетворительным может быть выявление признаков послеродовой депрессии у женщин при помощи Эдинбургской шкалы послеродовой депрессии?

Отчасти вопросы, касающиеся измерений, перекликаются с дескриптивно-сопоставительными и корреляционными вопросами, однако их спецификой является фокусирование внимания на характеристиках определенного измерительного инструмента, а не конструктора, измеряемого этим инструментом (Barker C. [et al.], 2002).

Задачи исследования. В окончательном виде сформулированные вопросы должны быть четкими и лаконичными, и в таком виде они должны быть отражены в виде задач исследования, которые предстоит решать в соответствии с целью исследования. Задачи обычно обозначаются в форме перечисления (изучить, описать, установить, выявить и т. п.). Формулировка задач исследования не допускает двусмысленности и неясности, поскольку описание их решения должно составить содержание соответствующих разделов работы. Важно сформулировать задачи исследования до начала процедуры исследования. Нередко начинающие авторы торопятся выбрать какую-то методику обследования и начать ее применять еще до того, как четко сформулируют вопросы и задачи, стоящие перед ними, причем выбор методики основан на ее известности и доступности. Такой подход неизбежно приводит к сужению круга вопросов, которые могут рассматриваться в работе, и ставит поставленные задачи в зависимость от выбранной методики, хотя последовательность должна быть обратной. Необходимо, чтобы выводы, сделанные автором по результатам исследования, содержали ответ на вопрос, заданный в пункте «цель исследования», и лучше всего придерживаться схемы: «одна задача — один раздел (глава) результатов — один вывод» (Марьянович А. Т., Князькин И. В., 2009).

Проверка гипотезы. Цель и задачи могут предполагать, что исследование будет проводиться как проверка гипотезы или как эксплоративное исследование. В идеальном случае теоретическая стадия анализа проблемы начинается с выдвижения и обоснования гипотезы, которая призвана дать пробное решение проблемы. Гипотеза должна дать шанс приблизиться к правильному решению поставленной проблемы или показать несостоятельность предпринятой попытки. Убедиться в этом можно лишь после проверки гипотезы. В научном исследовании гипотеза выступает как форма теоретического знания, содержащая предположение, сформулированное на основе ряда фактов, истинное значение которых неопре-

деленно и нуждается в доказательстве (Баскаков А. Я., Туленков Н. В., 2004). В процессе доказательства выдвинутых гипотез одни из них становятся истинной теорией, другие видоизменяются, уточняются и конкретизируются, третьи отбрасываются как заблуждения, если проверка дает отрицательный результат. Выдвижение новой гипотезы, как правило, опирается на результаты проверки предыдущей, даже в том случае, если эти результаты были отрицательными. Наличие четко сформулированной гипотезы делает исследование более сфокусированным. Кроме того, появляется возможность судить, насколько оправдывается гипотетическое предсказание по результатам исследования, какова его оценка с позиций статистической достоверности (вероятность нулевой гипотезы) (Howell D. C., 2013). Такой принцип является частью гипотетически-дедуктивного подхода к научному исследованию, который подчеркивает значимость теории и предыдущих исследований в генерировании проверяемой гипотезы. Суть подхода описана известным философом К. Поппером и состоит в следующем: из некоторой новой идеи с помощью логической дедукции выводятся следствия, которые можно назвать предсказаниями, в частности предсказания, которые легко проверяемы или непосредственно применимы. Затем выносятся некоторое решение относительно этих (и других) выводимых высказываний путем сравнения их с результатами практических применений и экспериментов. Если такое решение положительно, то есть если следствия оказываются приемлемыми или верифицированными, то теория может считаться выдержавшей проверку и у нас нет оснований отказываться от нее. Но если вынесенное решение отрицательное или, иначе говоря, если следствия оказались фальсифицированными, то фальсифицируется и сама теория, из которой они были логически выведены (Поппер К., 2004).

В связи с высказанными соображениями предпочтительно формулировать одновременно с задачами исследования и гипотезу исследования — совокупность предположений, допущений, которые уточняют путь достижения цели. Дальнейшее исследование выстраивается так, чтобы проверить и подтвердить истинность каждого положения гипотезы. Гипотеза должна быть достаточно простой и проверяемой. По логике она может иметь линейный (выдвигается и проверяется одно предположение) или разветвленный (состоит из нескольких предположений) вид. Для прогресса систематического мышления нужны некоторые общие рабочие гипотезы. Такие гипотезы направляют наблюдения, помогают оценить значение фактов различного типа и предписывают определенный метод. Даже неадекватная рабочая гипотеза, подтверждаемая хотя бы некоторыми фактами, все же лучше, чем ничего, так как она хоть как-то упорядочивает логические

познавательные процедуры (Уайтхед А., 1990). Гипотеза как форма теоретического знания должна отвечать общим требованиям, соблюдение которых необходимо не только для ее возникновения и обоснования, но и для построения любой научной гипотезы независимо от отрасли научного знания:

- соответствовать установленным в науке законам;
- согласовываться с фактическим материалом, на базе которого и для объяснения которого она выдвинута (она должна объяснить все имеющиеся достоверные факты);
- не содержать противоречий, которые запрещаются законами формальной логики;
- быть простой, не содержать лишнего, произвольных допущений, не вытекающих из необходимости познания объекта таким, каков он есть в действительности;
- должна допускать возможность ее подтверждения или опровержения: либо прямо — непосредственным наблюдением за явлениями, существование которых предполагается этой гипотезой, либо косвенно — выводением следствий из гипотезы и их последующей опытной проверкой (Баскаков А. Я., Туленков Н. В., 2004).

В ряде случаев задачи исследования могут быть сформулированы в форме открытых вопросов, что позволяет осуществить **эксплоративное исследование**, ориентированное на выявление новых фактов, когда отсутствует достаточная теоретическая база, позволяющая сформулировать обоснованную гипотезу, или когда исследователь не хочет ограничивать исследование рамками определенной гипотезы. Открытые вопросы можно считать наиболее приемлемыми в следующих случаях:

- если область исследования относительно новая и малоизученная, что делает преждевременным попытку ставить более конкретные вопросы;
- если область исследования неопределенна и противоречива, а в самой области исследования нет заметного продвижения вперед;
- если тема исследования связана с очень сложным феноменом, процессом или событием, требующим детального определения или описания (Barker C. [et al.], 2002).

В подобном случае исследование будет основано на индуктивном методе, который предполагает движение от частного к общему, т. е. от фактов к теории. Выводы, сделанные на основе анализа отдельных фактов и явлений, считаются справедливыми для всех других аналогичных фактов и явлений. Однако и в условиях эксплоративного исследования важно четко определить, что исследователь хочет изучать. Задачи эксплоративного исследования, как правило, состоят в описании определенного феномена. Они должны быть четко обо-

значены, чтобы ограничить область исследования рамками с целью сфокусировать усилия на сборе и анализе данных.

Общее теоретическое знание выводится исходя из закономерности, полученной в единичном, эмпирическом случае (от частного к общему). Строго говоря, понятие, выведенное индуктивным путем, является недоказуемым, а закон, полученный с помощью индукции, есть не что иное, как хорошая гипотеза, требующая проверки. Ниже сопоставлены логические построения и научные контексты гипотетически-дедуктивного и эксплоративного исследований (табл. 1, рис. 2).

Таблица 1

Проверка гипотезы и эксплоративное исследование
(Barker C. [et al.], 2002)

Характеристика	Проверка гипотезы	Эксплоративное исследование
Логика	Гипотетически-дедуктивная	Индуктивная
Научный контекст	Подтверждающий	Ориентированный на открытие
Пример 1. Лечение алкогольной зависимости	Является ли первый месяц лечения и трезвости психологически более трудным для пациента, чем шестой месяц?	Как протекает процесс лечения на протяжении первых шести месяцев?
Пример 2. Использование никотинового пластыря для борьбы с никотиновой зависимостью	Уменьшается ли у пациента, использующего никотиновый пластырь, желание закурить при виде провоцирующих стимулов (например, когда кто-то закуривает сигарету)?	Что испытывает пациент, использующий никотиновый пластырь для борьбы с курением (каковы его эмоции, когнитивные характеристики, ощущения)?

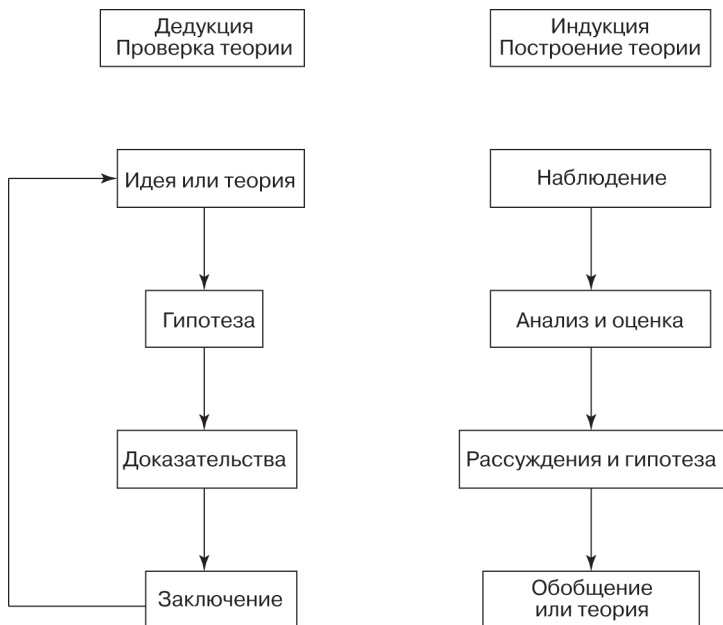


Рис. 2. Подходы к разработке теории (Newby P., 2014)

Вопросы для самоконтроля

1. В чем суть подготовительного этапа исследования?
2. Для чего требуется анализ литературы по проблеме исследования?
3. Каковы важнейшие электронные ресурсы научной литературы по клинической психологии?
4. Какими программными средствами можно воспользоваться для работы с библиографическими источниками?
5. Что такое гипотеза исследования? В чем различия гипотетически-дедуктивного и эксплоративно-индуктивного исследований?
6. Что такое цель и задачи исследования? На какие вопросы может отвечать исследование?

ДИЗАЙН ИССЛЕДОВАНИЯ

Дизайн исследования — это его общий план, описание того, как будет проводиться исследование. Выбор дизайна исследования предполагает ответы на вопросы, когда, где и с привлечением каких участников будут выполнены измерения тех или иных параметров. Прежде всего, важно определить его структуру и все аспекты сбора данных, включая подходы к измерению параметров и частоту проведения этих измерений (Кельмансон И. А., 2002; Мелихов О. Г., 2003; Hennekens С. Н. [et al.], 1987; Altman D. G., 1999; Barker С. [et al.], 2002; Bailar J. С., Hoaglin D. С., 2009) (рис. 3).

2.1. Описательные исследования

Предметом описательных исследований является изучение распространенности заболевания или иного состояния, включая поиск ответов на следующие вопросы:

- Какая часть популяции подвержена или не подвержена заболеванию (изучаемому состоянию)?
- В каком географическом регионе заболевание (изучаемое состояние) отмечается чаще или реже?
- Как меняется распространенность заболевания (изучаемого состояния) с течением времени? (Кельмансон И. А., 2002).

Подобные исследования представляют собой первый шаг по пути выявления факторов, влияющих на риск развития заболевания (патологического состояния). Используется информация, которую содержат данные переписи населения, регистры состояния здоровья, истории болезни, родов, развития ребенка и т. п. В связи с тем, что подобная информация, как правило, доступна, описательные исследования требуют меньших затрат, чем аналитические исследования.

Описательные исследования часто дают возможность сформулировать **рабочую гипотезу** относительно факторов риска, подлежащую проверке в ходе последующих аналитических исследований. Можно выделить три важнейших принципа формулировки такой гипотезы, основанных на индуктивном мышлении (Hennekens С. Н. [et al.], 1987; Newby P., 2014):

- **Принцип различий:** если частота заболевания существенно различается при внешних обстоятельствах двух типов, то весь

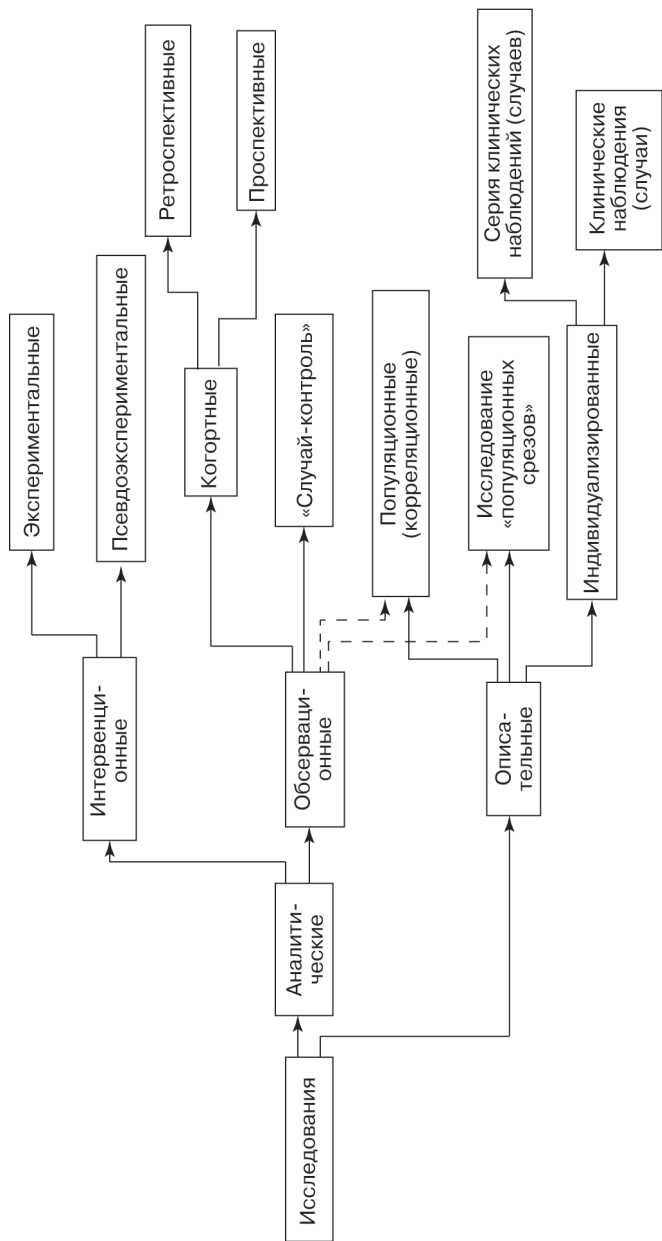


Рис. 3. Важнейшие варианты дизайна исследований

ма вероятно, что заболевание вызывается именно одним из факторов, обуславливающих различия этих внешних обстоятельств. Сопоставление распространенности заболеваний в отдельных регионах дает основания для такого подхода к формулировке рабочей гипотезы.

— **Принцип согласия:** если какой-либо специфический фактор часто определяется в тех ситуациях, в которых заболевание возникает с большой частотой, то данный фактор может быть причинно значимым.

— **Принцип сопряженной варибельности:** частота потенциального причинного фактора варьирует пропорционально частоте заболевания. Для формулирования гипотезы в этих случаях особое значение имеют корреляционные исследования.

Каков бы ни был принцип, лежащий в основе выработки рабочей гипотезы, существует несколько предварительных условий:

— рассмотрение общих описательных характеристик заболевания (патологического состояния), ведущими среди которых следует считать специфические черты пациентов, страдающих изучаемым заболеванием: сведения о распределении по возрасту и полу, особенности места и времени возникновения заболевания, возможно, учет ряда социально-демографических показателей, типологии личности и т. п.;

— поиск ответа на вопрос: где изучаемое заболевание (патологическое состояние) встречается с максимальной и минимальной частотой? Описание пространственного распределения случаев заболевания может иметь важнейшее значение для формулировки гипотезы относительно этиологии заболевания и факторов риска. Сопоставление географического распределения случаев заболевания возможно как в мировом масштабе с опорой на показатели, характерные для отдельных стран, так и в пределах одной страны или региона. Сопоставление показателей, характерных для отдельных стран, дает возможность изучить соотношение роли генетических и средовых факторов в генезе тех или иных заболеваний. Проводя сопоставления распространенности того или иного заболевания среди лиц определенной этнической группы, проживающих в различных регионах, можно косвенно судить о том, в какой мере средовые факторы модифицируют риск развития данного заболевания (патологического состояния).

Данные описательных исследований с целью определить места наибольшей концентрации случаев тех или иных заболеваний могут быть достаточно наглядно представлены на карте. Для этого предназначены прикладные статистические программы, например общедоступная программа EpiInfo (<https://www.cdc.gov/epiinfo>), один из модулей которой (EpiMap) позволяет выполнять подобные задачи. Однако при наличии признаков пространственной концентрации

случаев заболевания необходимо дальнейшее аналитическое исследование, направленное на подтверждение достоверности такой концентрации. В качестве примера представлены сведения о распространенности депрессивных расстройств в 2010 г. в различных регионах мира (Ferrari A. J. [et al.], 2013) (рис. 4, см. цв. вкл.).

Описательные исследования, затрагивающие временные аспекты, предполагают также поиск ответа на следующие вопросы:

- Когда заболевание возникает редко, а когда часто?
- Отличается ли частота заболевания в настоящее время от частоты заболевания в предшествовавшие годы?

Констатация возрастания частоты заболеваний и патологических состояний, развитие которых характеризуется коротким латентным периодом, с течением времени нередко позволяет выявить причинно значимый агент.

Пример

Приведем классические данные о тератогенном эффекте талидомида, когда в конце 1959 г. в Германии было зафиксировано единичное описание необычной формы врожденного порока развития конечностей и пальцев у ребенка, а вскоре появились новые аналогичные описания, и к сентябрю 1961 г. уже было зафиксировано необычайно высокое число случаев рождения детей с подобными дефектами. За 1957–1961 гг. частота такого рода врожденных пороков развития возросла в 200 раз по сравнению с предыдущими годами, что укладывалось в классическое определение эпидемии. Анализ фактов привел к тому, что подозрение пало на талидомид — снотворное средство, которое стало широко использоваться в Германии начиная с 1956 г. Последующие аналитические исследования действительно подтвердили, что женщины, которые сравнительно часто принимали его, рожали детей с грубыми пороками развития.

Пример

В конце 1980-х гг. статистика зафиксировала увеличение частоты синдрома внезапной смерти младенцев (СВСМ) в ряде промышленно развитых стран. В частности, в Норвегии пик случаев СВСМ составил 2,6 на 1000 родившихся живыми детей в 1988 г.; для сравнения: в начале 1970-х гг. данный показатель составлял около 0,7 на 1000. С конца 1970-х и до конца 1980-х гг. широкое распространение получила практика укладывать младенцев так, чтобы они спали на животе, в качестве аргументов указывались меньшая частота срыгивания и меньший риск аспирации пищевых масс во время сна у детей, особенно родившихся с малой массой тела и (или) недоношенными; более спокойный ночной сон ребенка и даже лучшие отдаленные показатели психического развития. Был выявлен отчетливый параллелизм между учащением практики ночного сна детей на животе и возрастанием случаев СВСМ. Именно этот факт послужил опорной точкой в поиске возможной ассоциации между двумя указанными обстоятельствами. В дальнейшем аналитические исследования выявили обоснованность такого предположения (Кельмансон И. А., 2002).

Имеются и иные варианты распределения случаев заболевания во времени, которые могут способствовать пониманию причинных факторов. Так, циклические изменения, например сезонные колебания, имеют большую ценность в описании острых заболеваний или заболеваний с коротким латентным периодом.

Выделяют три основных варианта описательных исследований:

— описания отдельных случаев (кейсов) или серии наблюдений (кейсов);

— корреляционные исследования, которые направлены на изучение распространенности заболевания в различных популяциях;

— кросс-секционные исследования, которые можно также называть исследованиями популяционных срезов.

Хотя по формальным признакам корреляционные и кросс-секционные исследования большинством авторов относятся к категории описательных, в ходе проведения таких исследований возможно осуществление определенной аналитической работы, о чем будет сказано ниже.

2.1.1. Описания клинических случаев и серии наблюдений

Описания клинических случаев и серии наблюдений заключаются в анализе клинико-психологических характеристик отдельно пациента или группы лиц со сходными клиническими симптомами или одинаковым диагнозом. Клинические описания отдельных случаев или серий наблюдений являются связующим звеном между клинической медициной, клинической психологией, с одной стороны, и клинической эпидемиологией — с другой. Нередко они имеют особую ценность для клиницистов и клинических психологов, так как позволяют выявить некоторые необычные аспекты течения заболевания или ситуации, предшествовавшей его развитию. Иногда это позволяет сформулировать рабочую гипотезу о природе возникшего заболевания или расстройства. Следует отметить, что описания клинических наблюдений составляют более трети публикаций в медицинских и клинико-психологических журналах, а некоторые журналы специализируются именно на публикации таких клинических кейсов (например, журнал *Clinical case reports*). Подобного рода исследования имеют принципиальное значение для распознавания новых заболеваний, особенно редко встречающихся, а также новых форм неблагоприятных исходов, связанных с рядом патогенных воздействий. Как известно, на описании отдельных клинических наблюдений основаны работы З. Фрейда. Описания серий клинических наблюдений содержат совокупность отдельных клинических случаев, нередко регистрируемых в течение достаточно короткого промежутка времени, и поз-

воляют на ранних стадиях зафиксировать эпидемическую вспышку того или иного заболевания. Существенным недостатком единичных клинических описаний является то, что они базируются на индивидуальных данных одного пациента. Выявляемые у него предположительные факторы риска заболевания могут быть лишь следствием случайного стечения обстоятельств. При описании серий наблюдений с большей вероятностью можно предположить достаточное число пациентов, позволяющее судить о встречаемости предполагаемого причинно значимого патогенного воздействия, однако интерпретация данных по-прежнему затруднена вследствие отсутствия адекватной контрольной группы наблюдений.

2.1.2. Исследование популяционных срезов

Исследование популяционного среза (*cross-sectional design, survey*) предполагает участие в исследовании в качестве субъектов всех членов популяции на время проведения исследования или представительную выборку этих членов популяции без учета их статуса по отношению к изучаемому воздействию и наличию или отсутствию изучаемого заболевания.

Исследование популяционного среза, направленное на изучение распространенности того или иного состояния (т. е. показателя преваленса), именуется **превалентным исследованием**. Обычно наличие или отсутствие интересующего исследователя патогенного воздействия оценивается одновременно с наличием или отсутствием заболевания (патологического состояния), и различные субпопуляции, сформированные в зависимости от наличия воздействия, сопоставляются между собой по показателю распространенности заболевания в них. В подобных исследованиях не ставятся задачи выявления этиологии.

Популяционные срезы изучаются в четко обозначенной целевой популяции (*target population*), соответствующей цели и задачам исследования. Прежде всего, такое исследование направлено на изучение распространенности того или иного заболевания (состояния, признака, феномена) в целевой популяции. При этом не исключается наличие или отсутствие у обследованных тех или иных факторов, способствующих, по мнению исследователя, повышению или снижению риска возникновения изучаемого заболевания.

Поскольку в исследованиях популяционных срезов наличие потенциального патогенного воздействия и наличие заболевания оцениваются одновременно, не возможно с уверенностью сказать, действительно ли анализируемое воздействие, рассматриваемое в качестве причинного фактора, предшествовало развитию заболевания

и способствовало его развитию, или, наоборот, повышенная частота этого фактора явилась следствием уже развившегося заболевания.

Необходимо четко определить временной промежуток, в течение которого проводится исследование популяционных срезов (например, календарный год). Возможна ориентация на какое-то определенное событие, связанное с естественным течением заболевания или естественным развитием процесса. При этом у отдельных обследованных лиц это событие может быть приурочено к разным моментам реального времени (например, к выписке ребенка из родильного дома, к медицинскому осмотру ребенка при поступлении в школу и т. п.). Результаты исследования во многом будут зависеть от этой выбранной временной точки отсчета. Сказанное можно проиллюстрировать следующим гипотетическим примером (рис. 5). Предположим, что проводится исследование с целью изучить ассоциацию между назначением детям витаминных препаратов и частотой аллергических реакций у них. Исходно 100 детей получали витаминные препараты, другие 100 детей – нет.

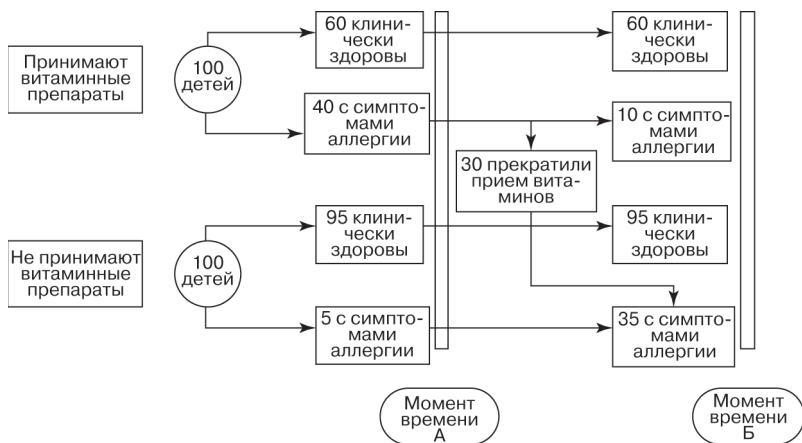


Рис. 5. Зависимость результатов исследования популяционного среза от выбора точки отсчета

Если исследование проводится в момент времени А, то можно выявить, что симптомы аллергии наблюдаются у 40 из 100 детей (40 %), получающих витаминные препараты, и у 5 из 100 (5 %), которым препараты не дают. Можно сделать заключение, что частота аллергических реакций в 8 раз выше среди детей, получающих витаминные препараты, чем среди детей, не получающих витамины. С учетом

естественного хода событий появление аллергических реакций у детей приведет к отказу родителей от назначения детям витаминных препаратов. Предположим, что 30 из 40 детей, имевших симптомы аллергии на фоне приема витаминных препаратов, откажутся от них в дальнейшем. Тогда при обследовании популяционного среза в момент времени Б аллергические реакции будут определяться у 10 из 70 детей (14,3 %), получающих в это время витаминные препараты, и у 35 из 130 детей (26,9 %), которым витаминов не дают. При неверной интерпретации этих данных можно сделать вывод о меньшей частоте симптомов аллергии среди детей, получающих витаминные препараты. На самом же деле снижение частоты приема витаминов будет следствием более частых аллергических реакций, а не причиной более редких аллергических реакций.

В целом в исследованиях популяционных срезов проблема причинно-следственных связей всегда является предметом дискуссий.

Исследования популяционных срезов всегда опираются на показатели распространенности изучаемого состояния (превалентные показатели), а не на частоту вновь регистрируемых случаев (инцидентные показатели). Значит, данные таких исследований будут отражать не только этиологические характеристики, определяющие возникновение новых случаев заболеваний, но и тенденцию к их хронизации, показатели выздоровления и выживаемости пациентов. Так, сопоставление распространенности в популяции (преваленса) детей-инвалидов в различных регионах мира будет отражать не только частоту возникновения инвалидизирующих заболеваний среди детского населения, но и выживаемость этих хронически больных детей: не представляется возможным судить о том, какая из двух тенденций будет ведущей, только по данным изучения популяционных срезов. Для изучения популяционных срезов можно использовать элементы аналитического исследования, т. е. предполагать выявление ассоциации между потенциальным патогенным воздействием и связанным с ним эффектом. Однако такой анализ может быть адекватным лишь в тех случаях, когда фактор, рассматриваемый в качестве потенциального патогенного воздействия, не претерпевает изменений в популяции с течением времени. В таком случае можно с уверенностью утверждать, что значения этого фактора, определяемые на момент обследования популяции, имели место и в начале развития заболевания. К числу таких факторов можно отнести лишь те, которые существуют в неизменном виде с момента рождения: цвет глаз, группу крови, антигены главного комплекса гистосовместимости и т. п. Чаще в ходе изучения популяционных срезов приходится иметь дело с факторами, потенциально меняющимися по мере развития заболевания, а иногда и непосредственно в связи с его естественным течением.

Большое значение имеет подход к определению обследуемого контингента. **Целевая популяция** (*target population*) — это та популяция, информацию о которой исследователю хотелось бы получить. Для ее определения исследователю необходимо ответить на ряд вопросов (Franklin S., Walker C., 2010):

— Что представляют собой те единицы наблюдения, из которых состоит популяция, и какие характеристики определяют эти единицы наблюдения (кто или что?)?

— Каково географическое расположение единиц наблюдения, формирующих популяцию (где?)?

— Каков референтный (временной) период наблюдения (когда?)?

Целевая популяция — это своего рода идеальная популяция, желательная для изучения, хотя этот идеал не всегда достижим. Определение целевой популяции может основываться на определенных концептуальных принципах даже при отсутствии списочного состава потенциальных участников исследования, например: все школьники младших классов. При этом нужно четко определить, что такое «младшие классы»? Будут ли относиться к числу потенциальных участников исследования дети, обучающиеся на дому, или речь идет только о тех, кто посещает школу?

Обследуемая популяция (*survey population*) — это популяция, которая фактически охвачена исследованием, и только на нее можно распространять результаты исследования. В идеале обследуемая популяция и целевая популяция должны быть очень похожи, хотя по разным причинам могут отличаться. Например, обследование целевой популяции может предполагать участие объектов, расположенных в отдаленных регионах, и их приходится исключить из последующего обследования.

После определения вопроса о целевой популяции нужно определить **рамки популяции** (*survey frame*) или **рамки выборки** (*sampling frame*), если проводится обследование выборки, которые позволяют идентифицировать единицы обследования, входящие в целевую популяцию, и обеспечить контакт с ними. Существуют два вида рамок: списочные и географические, если они не подошли, используются множественные рамки.

Списочные рамки представляют собой концептуальный или фактический список всех единиц, входящих в обследуемую популяцию. Например, концептуальным является список всех возможных пациентов, госпитализированных в определенный стационар, за определенный период времени, фактическим — реальный перечень потенциальных единиц обследования, почерпнутый из различных источников, например из медицинских регистров, адресных регистров и т. п.

Географические рамки — это список, в котором единицами наблюдения выступают географические территории.

Рамочная популяция (*frame population*) — обследуемая популяция, ограниченная рамками (списочными или географическими).

Популяция респондентов (*respondent population*) — те участники рамочной популяции, которые потенциально согласились бы принять участие в исследовании, если бы им было предложено. Члены рамочной популяции, которые могут быть отобраны для участия в исследовании, могут принять в нем участие, но могут и отказаться от участия в нем.

Соотношения указанных концептов представлено на рис. 6.

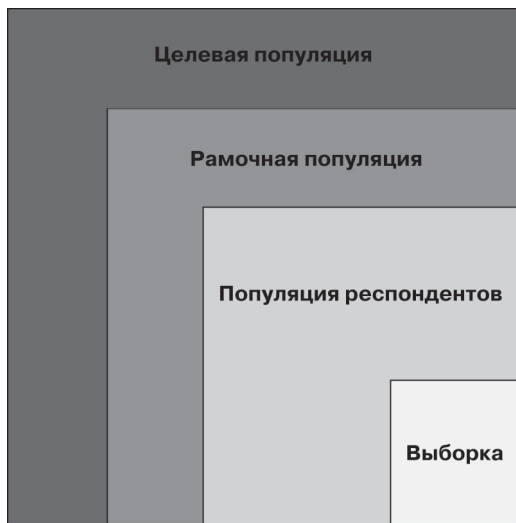


Рис. 6. Соотношение целевой, рамочной популяций, популяции респондентов и выборки (Leeuw E. D. [et al.], 2008)

Важнейшие подходы к исследованию популяционных срезов представлены на рис. 7. **Перепись** — самый эффективный способ проведения исследования, так как он предполагает обследование всей целевой популяции. Такой подход возможен лишь в ограниченных ситуациях (когда целевая популяция немногочисленна и ее члены доступны для обследования).

Выборка из целевой популяции чаще используется для исследования. Можно выделить два принципиальных подхода к формированию такой выборки.

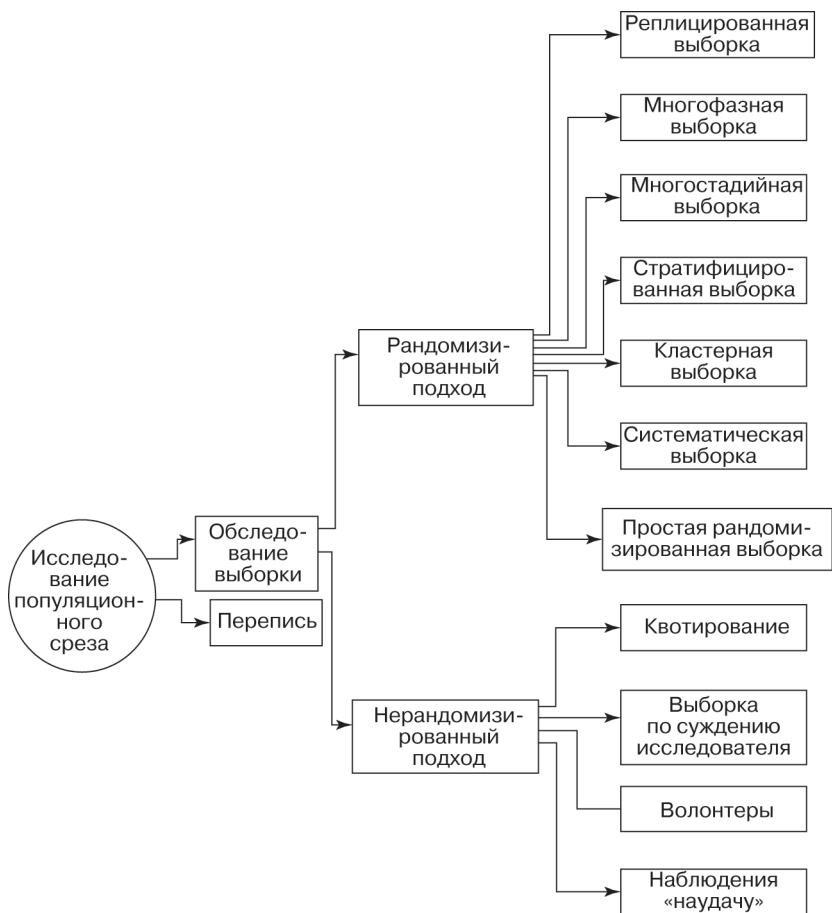


Рис. 7. Варианты проведения исследований популяционных срезов

Нерандомизированный (невероятностный) подход основан на отборе испытуемых, исходя из субъективных подходов исследователя. Это быстрый, относительно легкий и экономичный способ проведения исследования, который не предполагает установления рамок исследования. Основная проблема состоит в том, что нет уверенности, в какой степени можно обобщать результаты, полученные в ходе такого исследования, применительно к целевой популяции в целом. Следовательно, невероятностный подход к отбору обследуемых ин-

дивидуумов может приводить к большим систематическим ошибкам. Исследователь может с большой вероятностью отбирать наиболее доступных для него членов популяции, дружески расположенных к нему, и существенная часть членов целевой популяции просто не имеют шансов попасть в исследование, и эта часть популяции будет систематически отличаться от тех, кто был отобран для участия в исследовании. Такая ситуация может не только внести систематическое искажение в результаты исследования, но и сложно уменьшить кажущуюся вариацию популяции, поскольку отбираются лишь наиболее характерные индивидуумы и искусственно устраняются «экстремальные» наблюдения.

Вследствие наличия ошибки отбора испытуемых и (обычно) отсутствия обозначенных популяционных рамок, вероятность попадания отдельных индивидуумов в число обследованных не может быть рассчитана. Как аксиома принимается допущение, что обследуемая выборка является репрезентативной применительно к целевой популяции. При этом также предполагается, что характеристики обследованных лиц удовлетворяют какой-то определенной модели или равномерно или случайно распределены среди всех членов целевой популяции. Указанные допущения довольно рискованны, так как нет эффективных средств проверить их обоснованность. Нерандомизированные подходы можно использовать:

- для выдвижения определенной идеи;
- в качестве предварительного шага к разработке последующего исследования, основанного на рандомизированном подходе;
- в качестве последующего шага, который должен помочь лучше понять результаты исследования, ранее проводившегося на основе рандомизации (Franklin S., Walker C., 2010).

Так, нерандомизированное кросс-секционное исследование может использоваться для получения ценной информации относительно предпочтений, мотивации, особенностей поведения людей. Иногда такие нерандомизированные подходы могут быть единственным возможным вариантом, когда предполагается обследование именно волонтеров. Нерандомизированные исследования часто используются для формирования фокус-групп и подробного интервьюирования, например в ходе переписных исследований для выяснения того, насколько понятны сформулированные вопросы; если вопросы понятны членам отдельных специально отобранных фокус-групп, то можно рассчитывать, что они будут понятны всем членам популяции.

Достоинства нерандомизированного подхода

- Исследование можно провести быстро, а его выполнение технически удобно.

— Исследование не требует больших затрат. Интервьюер может в течение нескольких часов осуществить опрос обследуемых; кроме того, при нерандомизированном подходе обследуемые лица обычно широко не разбросаны географически, что уменьшает и транспортные расходы.

— Не требуется формирование рамок исследования.

— Такой подход может быть полезным при эксплоративном исследовании и дальнейшей разработке полномасштабного кросс-секционного исследования (Franklin S., Walker C., 2010).

Недостатки нерандомизированных исследований:

— Требуются очень сильные допущения относительно репрезентативности обследуемой выборки.

— Невозможно определить вероятность, с которой тот или иной индивидум оказался в числе обследуемых лиц (Franklin S., Walker C., 2010).

Рассмотрим отдельные варианты нерандомизированных подходов.

Метод формирования выборки «наудачу» (стихийно) (*haphazard sampling*). Единицы наблюдений отбираются без какой-то заранее заданной цели и плана, например интервьюирование «первого встречного» человека на улице. Предполагается, что вся популяция однородна (гомогенна), а следовательно, все члены популяции похожи друг на друга, поэтому случайным образом могут быть отобраны в качестве единиц исследования.

Метод формирования выборки из числа волонтеров (*volunteer sampling*). Респондентами становятся волонтеры. Обычно их необходимо предварительно проверить на предмет соответствия цели исследования (например, наличие определенного заболевания, которое подвергается изучению). Такой метод сопряжен с высоким риском ошибки отбора, однако в ряде случаев он является необходимым. Например, по этическим соображениям пациентам с определенными заболеваниями может быть оказана медицинская и психологическая помощь. Обследование выборки из числа волонтеров может выступать в качестве предварительного этапа для формирования фокус-группы, которую в дальнейшем изучат более подробно. Следует избегать попыток обобщения полученных результатов в отношении популяции в целом.

Формирование выборки по суждению исследователя (*judgement sampling*). Выборка формируется на основе ранее сформированных представлений исследователя о составе популяции и поведении обследуемых. Эксперт решает, что собой представляет репрезентативная выборка и кто из членов популяции может стать участником исследования и войти в выборку. Формирование выборки по

суждению исследователя сопряжено с высокой вероятностью ошибки исследователя, а результаты такого исследования могут оказаться еще более искаженными, чем при формировании выборки «наудачу». В то же время такое исследование может быть полезным при постановке эксплоративных вопросов, например при формировании фокус-группы для испытания опросника.

Формирование выборки методом квотирования (*quota sampling*). Набор испытуемых проводится до тех пор, пока не будет достигнуто запланированное число единиц наблюдений (квот) различных субпопуляций. Формирование выборки методом квотирования позволяет набрать нужное число наблюдений различных субпопуляций. Квоты могут быть основаны на пропорциональном принципе. Например, если в целевой популяции имеется 100 мужчин и 100 женщин, а в выборку предполагается включить 20 человек, то квотирование) предполагает обследование 10 мужчин и 10 женщин. Этот метод наиболее распространен и более предпочтителен по сравнению с другими нерандомизированными методами формирования выборки, так как предполагает активное включение членов различных субпопуляций в исследование. Метод квотирования напоминает метод стратифицированного формирования выборки, поскольку сходные по каким-то признакам единицы наблюдения сводятся в одну группу. Если при вероятностном (рандомизированном) подходе такие единицы наблюдений отбираются по принципу рандомизации, то при квотировании решение вопроса о том, кого включить в исследование, остается за самим исследователем. Испытуемые, которые отказываются от участия в исследовании, часто просто замещаются другими участниками, и в результате общее необходимое число участников будет обеспечено. По сути, это приводит к игнорированию ошибки, связанной с отказом от участия в исследовании (*non-response bias*). Формирование выборки методом квотирования обычно требует меньше ресурсов, чем стратифицированное исследование, и его легче осуществить. Однако такой дизайн сопряжен с высоким риском ошибки отбора (*selection bias*) и, как в случае иных нерандомизированных подходов, требует значительного допущения: отобранные участники исследования сходны с теми, кто не вошел в исследование. Такие допущения редко соответствуют истине.

Вероятностный (рандомизированный) подход направлен на устранение систематической ошибки, которая описана выше. Формирование выборки является методологически более строгим. Преимущество состоит в том, что результаты, полученные при обследовании выборки, позволяют делать суждения в отношении популяции в целом. Вероятностный принцип формирования выборки позволяет избежать ошибки отбора за счет случайного (рандомизи-

рованного) отбора единиц обследования из популяции (при помощи компьютера или таблицы случайных чисел). В данном термине слово «случайный» не означает «произвольный». «Рандомизированное» означает, что отбор испытуемых свободен от систематической ошибки, — он основан на случае. Отбор испытуемых никогда не отдается «на откуп» интервьюеру, который субъективно решает, кого включить в выборку. При вероятностном подходе к формированию выборки, во-первых, все единицы обследования случайно (рандомизированно) отображены для участия в исследовании; во-вторых, все члены популяции имеют ненулевую вероятность быть включенными в выборку, и эти вероятности могут быть рассчитаны. Не обязательно, чтобы все единицы обследования имели равную вероятность быть включенными в выборку, и в сложных исследованиях такая вероятность варьирует у отдельных единиц.

Существуют различные варианты вероятностного дизайна.

Метод простой случайной выборки (simple random sampling) можно назвать базовым подходом. Он оптимален, если цель исследования состоит в оценке общепопуляционных характеристик, а стратификация невозможна или неприемлема.

Кластерное формирование выборки часто используется, если позволяют ресурсы.

Стратифицированное формирование выборки обычно применяется, если желательно изучение показателей в субпопуляциях (например, в отдельных регионах, возрастных группах, социальных стратах и т. п.).

Остальные варианты формирования выборки являются еще более сложными и обычно используют дополнительную информацию для обозначения рамок исследования.

Основное достоинство вероятностного подхода к формированию выборки: поскольку каждая единица обследования отбирается рандомизированно и вероятность включения в исследование каждой такой единицы может быть просчитана, то можно вычислить и надежные измеряемые показатели, а также величину ошибки этих показателей. Таким образом, можно судить о популяции в целом. Фактически можно делать выводы, касающиеся популяции в целом, исходя из сведений, полученных на относительно небольшой выборке.

Основным недостатком является относительная сложность дизайна, его реализация требует большего времени, а само исследование обходится дороже. Исследования с применением вероятностного подхода обычно предполагают и более широкий географический охват, существенно большую численность целевой популяции, что также создает трудности в реализации исследования.

Простая случайная выборка (*simple random sampling, SRS*) представляет собой метод одношагового отбора, обеспечивающий ситуацию, при которой имеется равная вероятность формирования любой выборки численностью n . Как следствие, каждая единица наблюдения, включенная в выборку, имеет равную вероятность π попадания в нее:

$$\pi = n/N.$$

Единицы наблюдения отбираются с замещением или без замещения. Отбор с замещением предполагает, что единица наблюдения может отбираться в выборку более одного раза. Отбор без замещения означает, что единица наблюдения отбирается только один раз. Простая случайная выборка с замещением (*simple random sampling with replacement, SRSWR*) и простая случайная выборка без замещения (*simple random sampling without replacement, SRSWOR*) практически идентичны, если численность выборки — очень малая доля общей популяции. В этом случае вероятность повторного выпадения одной и той же единицы наблюдения весьма мала. В целом формирование выборки без замещения дает более точные результаты и более удобно с практической точки зрения, поэтому именно такой вариант и будет рассматриваться в ходе дальнейшего изложения.

Представим, что популяция состоит из 5 единиц (5 пациентов), из которых нужно выбрать 3 на основании процедуры формирования простой случайной выборки без замещения. Всем пациентам присваиваются индивидуальные номера от 1 до 5, и популяция рассматривается как множество: {1, 2, 3, 4, 5}. Можно сформировать 10 различных вариантов подмножеств (выборок) численностью 3 человека: {1, 2, 3}, {1, 2, 4}, {1, 2, 5}, {1, 3, 4}, {1, 3, 5}, {1, 4, 5}, {2, 3, 4}, {2, 3, 5}, {2, 4, 5}, {3, 4, 5}. Каждая выборка имеет равную вероятность быть сформированной, и у каждого индивида существует вероятность попасть в 6 из 10 возможных выборок и вероятность отбора $\pi = n/N = 3/5$.

Для формирования простой случайной выборки следует обозначить полные рамки исследования. Все единицы наблюдений, входящие в списочные рамки, получают номера от 1 до N . Затем из этого списочного состава на основе случайного отбора отбирают заранее определенное число наблюдений n , которые и должны сформировать выборку. Для этого используют либо таблицу случайных чисел, либо генерацию случайных чисел на компьютере (большая часть прикладных статистических программ позволяет сгенерировать такую последовательность случайных чисел). Первые n наблюдений, имеющие соответствующие номера, и составят выборку.

Пример

Планируется обследование студентов института. Имеется возможность составить список этих студентов (списочные рамки) на основе имеющихся источников. Список включает в себя $N = 1530$ студентов, а для проведения

исследования требуется $n = 90$ студентов (выборка), которых необходимо отобрать методом простого случайного отбора. Такой отбор можно осуществить при помощи таблицы случайных чисел. На первом этапе нужно выбрать четырехзначное число (поскольку у нас 1530 студентов — тоже четырехзначное число). Его можно выбрать в любом месте таблицы (первые четыре разряда числа, приведенного в ней), а затем следует перемещаться по таблице в любом направлении. При этом нужно фиксировать те первые 90 четырехзначных числа, которые встретились по ходу перемещения и значения которых не превышали 1530. Предположим, что мы начали с ряда 01 и колонки 85—89 (табл. 2). Двигаясь вниз по колонке, мы выделим числа: 189, 256, 984, 744, 1441, 617 и т. д. Такой отбор чисел нужно продолжать до тех пор, пока не будут отобраны 90 различных чисел. В результате мы получим номера тех 90 студентов, которые должны будут войти в исследование и сформируют выборку. Поскольку речь идет о процедуре без замещения, то в тех случаях, когда одно и то же число при продвижении по таблице встречается повторно, его пропускают. По сходному алгоритму выполняется компьютерное генерирование случайных чисел.

Преимущества простой случайной выборки:

- метод самый простой из всех вероятностных методов;
- метод достаточно хорошо разработан теоретически, существуют принятые подходы к расчету численности выборки, оценки изучаемых показателей и их вариации;
- для формирования выборки не требуется дополнительной информации, касающейся рамок обследуемой популяции, достаточно полного списка всех членов обследуемой популяции и их контактных данных;
- метод не требует применения специальных технических средств (Franklin S., Walker C., 2010).

Недостатки простой случайной выборки:

- Метод не предполагает использования дополнительной информации, даже если она доступна и может быть использована. Как следствие, полученные оценки оказываются статистически менее эффективными по сравнению с теми, которые могут быть достигнуты при использовании других подходов к формированию выборки.
- Метод чреват большой трудоемкостью исследования и большими материальными затратами, поскольку члены сформированной выборки могут быть достаточно широко разбросаны географически.
- Не исключена возможность формирования «неудачной» выборки. Поскольку все выборки численностью n наблюдений имеют равные шансы быть сформированными, есть вероятность отбора таких участников целевой популяции, которые недостаточно сильно отличаются друг от друга (в том числе и отдалены географически) и, как следствие, могут недостаточно хорошо представлять популяцию в целом (Franklin S., Walker C., 2010).

Фрагмент таблицы случайных чисел
(Franklin S., Walker C., 2010)

	50–54	55–59	60–64	65–69	70–74	75–79	80–84	85–89	90–94	95–99
00	59311	58030	52098	87024	14194	82848	04190	96574	90464	29065
01	98567	76364	77204	27062	53402	96621	43918	01896	83991	51141
02	10363	97518	51400	98342	24830	61891	27101	37855	06235	33516
03	86852	19558	64432	99612	53537	59798	32803	67708	15297	28612
04	11258	24591	36863	31721	81305	94335	34936	02566	80972	08188
05	95068	84628	35911	33020	70659	80428	39936	31855	34334	64865
06	54463	47437	73804	36239	18739	72824	83671	39892	60518	37092
07	16874	62677	57412	31389	56869	62233	80827	73917	82402	84420
08	92484	63157	76593	03205	84869	72389	96363	52887	01087	66591
09	15669	56689	35682	53256	62300	81872	35213	09840	34471	74441
10	99116	75486	84989	23476	52967	67104	39495	39100	17217	74073
11	15696	10703	65178	90637	63110	17622	53988	71087	84148	11670
12	97720	15369	51269	69620	03388	13699	33423	67453	43269	56720
13	11666	13841	71681	98000	35979	39719	81899	07449	47985	46967
14	71628	73130	78783	75691	41632	09847	61547	18707	85489	69944
15	40501	51089	99943	91843	41995	88931	73631	69361	05375	15417
16	22518	55576	98215	82068	10798	82611	36584	67466	69377	40054
17	75112	30485	62173	02132	14878	92879	22281	16783	86352	00077
18	08327	02671	98191	84342	90813	49268	95441	15496	20168	09271
19	60251	45548	02146	05597	48228	81366	34598	72856	66762	17002
20	57430	82270	10421	00540	43648	75888	66049	21511	47676	33444
21	73528	39559	34434	88596	54086	71693	43132	14414	79949	85193
22	25991	65959	70769	64721	86413	33475	42740	06175	82758	66248
23	78388	16638	09134	59980	63806	48472	39318	35434	24057	74739
24	12477	09965	96657	57994	59439	76330	24596	77515	09577	91871
...										
45	12900	71775	29845	60774	94924	21810	38636	33717	67598	82521
46	75086	23537	49639	33595	31484	97588	28617	17979	78749	35234
47	99445	51434	29181	09993	38190	42553	68922	52125	91077	40197
48	26075	31671	45386	36583	93459	48599	52022	41330	60650	91321
49	13636	93596	23377	51133	95126	61496	42474	45141	46660	42338

Схематически подход к формированию простой случайной выборки представлен на рис. 8.

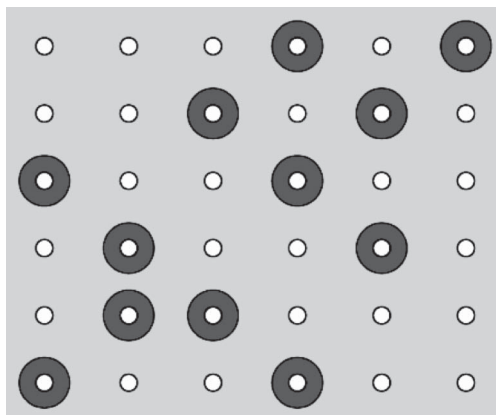


Рис. 8. Формирование простой случайной выборки численностью 12 наблюдений (Franklin S., Walker C., 2010)

Систематический отбор выборки (*systematic sampling, SYS*).

Единицы наблюдения отбираются в регулярных интервалах в популяции (рис. 9). К систематической выборке нередко прибегают тогда, когда хотелось бы использовать простую случайную выборку, но отсутствует списочный состав членов популяции или такой список имеется, но он уже исходно составлен в случайном порядке, и при этом легче организовать отбор систематической выборки, а не простой случайной. В случае систематической выборки необходимо определить интервал отбора единиц наблюдения и начальную случайно выбранную точку «старта». Если используется списочный состав членов популяции и число членов популяции равно N , число необходимых членов выборки равно n , и при этом отбирается каждый k -й объект, то $k = N/n$. Случайно выбранная точка «старта» r представляет собой число в диапазоне от 1 до k включительно. Выбранные единицы наблюдения будут следующими: $r, r + k, r + 2k, \dots, r + (n - 1)k$. Каждый объект наблюдения имеет вероятность попадания в выборку π , равную n/N . Однако, в отличие от простой случайной выборки, не каждая комбинация из n единиц имеет равные шансы быть отобранными: возможен только отбор единиц наблюдения, отделенных друг от друга значением k .

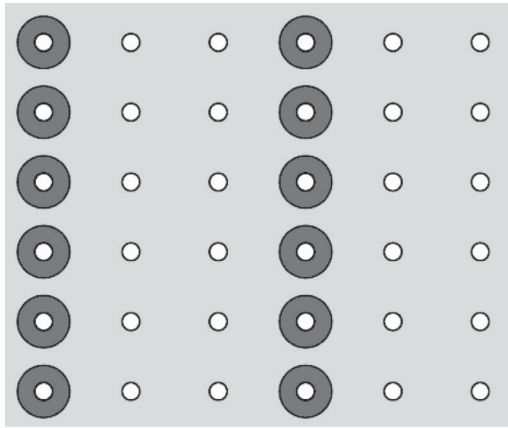


Рис. 9. Схема систематического формирования выборки. Общее число членов популяции $N = 36$, $n = 12$, $k = 3$ (Franklin S., Walker C., 2010)

Пример

Популяция состоит из 54 единиц наблюдения, и требуется формирование выборки численностью $n = 9$ единиц. Интервал для отбора k будет составлять $54/9 = 6$. Далее требуется выбор случайного числа в диапазоне от 1 до $k = 6$. Предположим, что таким числом оказалось 2. Тогда выбранными единицами наблюдения будут: 2, 8, 14, 20, 26, 32, 38, 44, 50. При интервале отбора 6 и размере популяции 54 единицы возможно формирование лишь 6 вариантов выборок.

Достоинства систематического формирования выборки:

— Метод является аналогом простой случайной выборки в тех случаях, когда нет четких рамок целевой популяции.

— Метод не требует дополнительной информации о рамках популяции.

— Метод может обеспечить формирование выборки с лучшей дисперсией, чем простая случайная выборка (в зависимости от интервала формирования выборки и того, как отсортирован список).

— Метод основан на разработанных теоретических предпосылках, как и метод формирования простой случайной выборки, поэтому параметры такой выборки могут быть легко рассчитаны.

— Метод проще, чем метод простой случайной выборки, так как требуется лишь одно случайное число.

Недостатки систематического формирования выборки:

— Метод может привести к формированию неудачной выборки, если выбранный интервал соответствует какой-то определенной собственной периодичности, существующей в популяции.

— Метод не использует какой-то дополнительной доступной информации относительно рамок популяции и поэтому может приводить к неэффективной стратегии формирования выборки.

— Окончательный размер выборки заранее не известен, если исследователь ориентируется лишь на предполагаемые рамки изучаемой популяции.

Кластерное формирование выборки (*clustersampling*) представляет собой процесс рандомизированного отбора целых групп единиц наблюдений (кластеров) в рамках изучаемой популяции (рис. 10). Обычно это менее эффективный подход к формированию выборки по сравнению с простой случайной выборкой. Формирование выборки методом кластеров снижает расходы на сбор информации, особенно если популяция широко разбросана и предполагается персональное интервьюирование членов выборки. Индивидуальный отбор испытуемых в популяции, не всегда практически оправдан. Иногда значительно проще формировать кластерные группы обследуемых (например, отдельных больниц, домов и т. п.). Наконец, такой подход позволяет дать оценку самим кластерам. Кластерный метод формирования выборки осуществляется в два этапа:

— популяция группируется в кластеры (это могут быть естественные кластеры, например: отдельные больницы, родильные дома, школы и т. п.);

— выполняются отбор кластеров и обследование всех единиц наблюдений, входящих в кластеры.

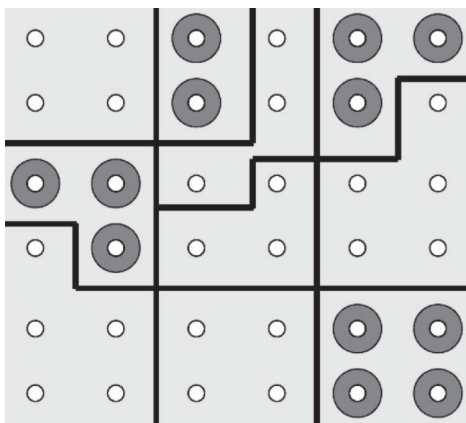


Рис. 10. Схема кластерного метода формирования выборки. Выборка представлена 4 кластерами (Franklin S., Walker C., 2010)

Для обеспечения эффективности кластерного метода формирования выборки нужно быть уверенным в том, что сами по себе единицы наблюдений, входящие в кластеры, максимально разнообразны. Если единицы наблюдений в пределах кластера однотипны, они все дают однотипную информацию, а значит, достаточно обследования одного участника кластера. К сожалению, единицы наблюдений в пределах кластеров часто имеют сходные характеристики и потому более гомогенны, чем случайно отобранные из общей популяции.

Достоинства кластерного формирования выборки:

— Метод может существенно уменьшить материальные затраты, особенно при широком разбросе членов популяции.

— Метод легче применять при естественной кластеризации населения (например, в пределах домовладения, школы) или при исследовании концептуальной популяции, например лиц, госпитализируемых в стационар в течение определенного периода времени.

— Метод позволяет оценить характеристики кластеров как таковых: например, при обследовании школьников можно одновременно изучить, каково общее число учителей в школе (если в качестве кластера выступают школы).

— Метод может быть статистически более эффективным, чем метод формирования простой случайной выборки, если единицы наблюдений, входящие в кластеры, гетерогенны (различны) по отношению к изучаемым показателям, а сами кластеры гомогенны (сходны). Однако на практике эти условия редко выполняются (Franklin S., Walker C., 2010).

Недостатки кластерного формирования выборки:

— Метод может быть менее эффективным статистически, чем метод простой случайной выборки, если единицы наблюдений в пределах кластеров гомогенны в отношении изучаемых показателей. Для того чтобы уменьшить влияние этого фактора, следует увеличить число кластеров.

— Окончательный размер выборки часто точно не известен заранее, так как нельзя сказать, сколько единиц наблюдений находится в пределах кластера до окончания исследования.

— Организация исследования может оказаться сложнее, чем в других методах.

Стратифицированное формирование выборки (*stratified sampling, STR*) предполагает, что популяция делится на гомогенные, взаимоисключающие группы, именуемые стратами, а затем в пределах каждой страты независимо отбираются объекты наблюдения (рис. 11). Формирование выборки в пределах страты проводится по одному из описанных принципов, в частности простой случайной выборки или систематической выборки, а также в соответствии с более

сложными приемами, например: пропорциональной выборки, кластерного формирования выборки или многофазного подхода к формированию выборки.

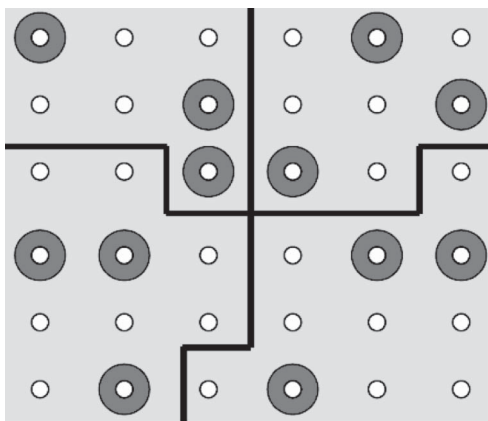


Рис. 11. Схема стратифицированного подхода к формированию выборки. Имеются 4 страты и 3 единицы наблюдения в каждой страте (Franklin S., Walker C., 2010)

Сочетание стратифицированного и кластерного формирования выборки реализуется в форме **стратифицированного кластерного подхода**. Например, при использовании кластерного метода формирования выборки часто на первом этапе проводят стратификацию, а затем в рамках каждой страты отбирают кластеры, которые обследуются в дальнейшем. Популяция может быть стратифицирована по любому признаку, который может быть учтен у всех единиц обследования в рамках целевой популяции еще до начала исследования. Например, таким признаком может быть просто домашний адрес (стратификация по географическому принципу: району, городу и т. п.) или сведения о доходе семьи (стратификация по уровню доходов). Чаще всего в качестве признаков для стратификации используются возраст, пол, географические характеристики, уровень доходов, жилищно-бытовые условия, образование и т. п.

Можно назвать основные доводы в пользу проведения стратификации:

— Стратегия стратификации более эффективна, чем при простой случайной выборке или систематическом отборе выборки.

— Стратификация позволяет обеспечить адекватную численность выборки в пределах тех специфических категорий, которые представляют интерес для исследователя.

— Стратификация отчасти защищает от формирования неудачной выборки: при заданной численности выборки стратификация приводит к снижению ошибки отбора.

Статистическая эффективность стратификации возрастает тогда, когда имеется сильная гомогенность единиц наблюдения в пределах каждой страты, а сами страты максимально отличаются друг от друга по тому признаку, который интересует исследователя. Как правило, это условие выполняется, когда признаки, по которым осуществляется стратификация, коррелируют с теми признаками, которые изучаются в ходе исследования. Стратификация также весьма эффективна, когда распределение анализируемого показателя в популяции отклоняется от нормального распределения (имеет выраженную асимметрию). В таких ситуациях несколько единиц наблюдений, относящихся к крайним точкам распределения, могут сильно повлиять на полученные оценки. Для того чтобы такого не случилось, эти экстремальные наблюдения должны войти в отдельную страту. В частности, распределение уровня доходов семьи в популяции будет характеризоваться асимметрией: очевидно, что в популяции число семей с очень высокими доходами будет меньшим, чем число семей со средним и низким доходом. Если в простую случайную выборку попадут несколько таких семей с высокими доходами, то они могут сильно повлиять на полученные оценки, прежде всего в тех случаях, когда эти оценки прямо или косвенно связаны с уровнем дохода (например, при изучении организации быта ребенка, его обеспеченности игрушками и т. п.). Такую ошибку предотвращает стратификационный подход к формированию выборки (со стратификацией по уровню доходов). В то же время, если исследование направлено, например, на изучение темперамента ребенка, то проводить стратификацию по уровню доходов нет смысла. Иными словами, стратификация, эффективная при изучении одного показателя, может оказаться неэффективной при изучении другого. Таким образом, признаки, по которым осуществляется стратификация, следует выбирать так, чтобы они коррелировали с наиболее важными показателями, изучаемыми в ходе исследования.

Важным соображением, которое учитывается при осуществлении стратификации, является обеспечение адекватной численности выборок в пределах известных доменов интереса (*domains of interest*). Кросс-секционное исследование проводится с общей целью — получение каких-то обобщенных характеристик. Однако в дополнение к общей характеристике исследователь может быть заинтересован

изучить данный показатель в пределах определенной подгруппы популяции, именуемой областью (доменом). Например, общая цель — узнать, какова распространенность высокого уровня тревоги у младших школьников. В пределах подгруппы цель — узнать распространенность высокого уровня тревоги отдельно среди младших школьников-мальчиков и девочек. Оценки показателей в подгруппах именуется оценкой домена. Если в ходе исследования предполагается оценка домена, то следует заранее предусмотреть, чтобы в пределах каждого домена имелось достаточное число наблюдений. Когда имеется информация о рамках популяции, простейшим способом обеспечить такое условие является строгое соответствие между стратами и доменами интереса.

Стратификация является защитой от формирования выборки с большим числом «экстремальных» представителей популяции. Формирование простой случайной выборки всецело зависит от случая. Если ее применяют к пациентам, то может оказаться, что в число обследованных вошли преимущественно мужчины или женщины; чтобы избежать такой крайности, следует провести стратифицированный отбор со стратификацией по полу. При этом предполагается, что в рамках целевой популяции имеется информация о поле пациентов.

Число единиц наблюдений, отбираемых для выборки в пределах каждой страты, может быть либо равным, либо пропорциональным численности каждой страты. В первом случае единицы наблюдения могут иметь различную вероятность попадания в обследуемую выборку. Рассмотрим в качестве примера популяцию численностью $N = 1000$ единиц, которую стратифицируют в две группы: одна страта включает в себя $N_1 = 250$ единиц наблюдений, другая — $N_2 = 750$ единиц. Предположим, что используется простая случайная выборка, чтобы случайно выбрать по 50 единиц наблюдений из первой и второй страт ($n_1 = n_2 = 50$). При этом общая численность выборки составит 100 наблюдений. Однако в первой страте вероятность попадания в обследуемую выборку составит: $\pi_1 = 50/250 = 0,2$; во второй: $\pi_2 = 50/750 = 0,067$.

Достоинства стратифицированного формирования выборки:

— Метод позволяет повысить точность оценок, обеспечивая более эффективную стратегию формирования выборки. Как следствие, решение задач исследования может быть обеспечено при меньшей численности выборки.

— Метод позволяет гарантировать, что важные подгруппы, если они определены как страты, будут достаточно представлены в выборке, что позволит дать статистически эффективные оценки доменов.

— Метод позволяет более эффективно проводить исследование в оперативном и административном плане.

— Метод позволяет избежать формирования «неудачной» выборки.

— Метод позволяет использовать различные выборочные рамки (*sampling frames*) и применять разные процедуры формирования выборки в отношении отдельных страт (например, простая случайная выборка в одной страте и вероятностная выборка, пропорциональная объему в другой) (Franklin S., Walker C., 2010).

Недостатки стратифицированного формирования выборки:

— Метод требует того, чтобы в рамки формирования выборки было включено много дополнительной и качественной информации, характеризующей все единицы наблюдений, ограниченные этими рамками. Это необходимо для осуществления стратификации.

— Формирование рамок исследования требует максимальных затрат по сравнению с другими методами формирования выборки.

— Стратегия формирования выборки может оказаться статистически менее эффективной, чем простая случайная выборка, если речь идет об изучении показателей, не коррелирующих с переменными, лежащими в основе стратификации.

— Оценки при этом методе более сложные, чем в случае простой случайной выборки или систематического отбора выборки.

Многостадийное формирование выборки (*multi-stage sampling*) представляет собой процесс формирования выборки, проходящий две или более последовательные стадии. Единицы наблюдения, отбираемые на первой стадии, именуется *первичными единицами выборки* (*primary sampling units, PSU*), единицы, отбираемые на второй стадии, — *единицами второй стадии* (*second stage units, SSU*) и т. д. (рис. 12). На каждой стадии единицы наблюдения различаются по структуре и расположены в иерархическом порядке (например, город — район — квартал — отдельный дом — люди, проживающие в доме). При двустадийном формировании выборки единицами второй стадии чаще всего выступают отдельные члены популяции.

При многостадийном формировании выборки чаще всего применяется двустадийный кластерный подход. На первой стадии используются географические рамки для отбора отдельных территорий, которые войдут в исследование, — первичных единиц выборки), к которым будет применен систематический подход к формированию выборки для отбора отдельных домов в пределах этих территорий — единиц второй стадии. При двустадийном кластерном подходе исследование будет проводиться лишь в некоторых единицах наблюдения. Если соседние единицы наблюдения в пределах одного кластера потенциально похожи друг на друга, то с точки зрения статистики более разумно анализировать небольшое число единиц второй стадии, выбранных из многих первичных единиц выборки, а не наоборот. В случае одностадийного кластерного подхода каждая единица наблюдения, вошедшая в кластер, должна была быть объектом исследования.

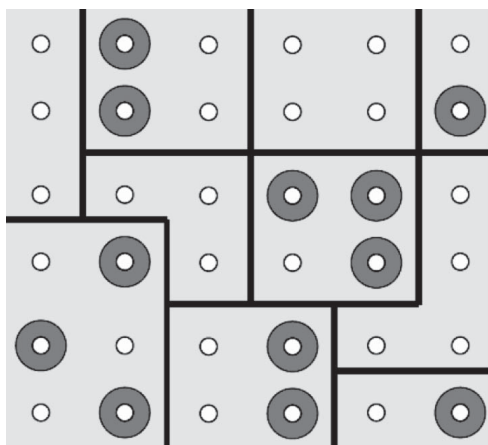


Рис. 12. Схема многостадийного формирования выборки. Отобрано 6 первичных единиц выборки с от 1 до 3 единиц второй стадии (Franklin S., Walker C., 2010)

Многостадийные наблюдения могут предполагать неограниченное число стадий формирования выборки. Однако сложность дизайна и оценки результатов возрастает по мере увеличения числа стадий, поэтому, как правило, число стадий ограничивают двумя или тремя. Следует отметить, что рамки первой стадии, как правило, достаточно стабильны. Например, если рамки имеют географическое выражение, то они редко меняются за короткий срок. Однако рамки второй и последующих стадий обычно менее стабильны (как правило, это списочные рамки, которые могут меняться с течением времени).

Принципиальным является то, что на каждой стадии многостадийного формирования выборки процесс осуществляется при помощи любого известного подхода, что делает многостадийный принцип весьма гибким.

Хотя географические рамки чаще всего используются на первой стадии многостадийного подхода к формированию выборки, соблюдение этого правила не является обязательным. Так, первичной единицей выборки может быть временной промежуток. Например, многостадийное исследование может предполагать изучение пациентов, госпитализированных в определенную многопрофильную больницу. В этом случае первичной единицей выборки может выступать определенный период времени (например, конкретные дни в течение месяца), а единицами второй стадии — отдельные пациенты, госпитализированные в эти дни и отобранные методом простой случайной выборки.

Достоинства многостадийного формирования выборки

— Метод позволяет обеспечить статистически более эффективную стратегию формирования выборки по сравнению с простым кластерным методом в тех случаях, когда кластеры гомогенны.

— Метод позволяет существенно уменьшить расходы на проведение исследования.

— Нет необходимости иметь списочные рамки всей популяции. Рамки нужно формировать на каждой стадии отбора популяции (Franklin S., Walker C., 2010).

Недостатки многостадийного формирования выборки:

— Метод не является столь же эффективным, сколь метод формирования простой случайной выборки.

— Окончательный размер выборки заранее не известен в случае многостадийного подхода, поскольку нет информации, сколько единиц наблюдения находится в пределах каждого кластера (размер выборки можно контролировать тогда, когда из каждого кластера планируется отбирать на второй стадии заранее зафиксированное число единиц наблюдения).

— Многостадийное обследование труднее организовать, чем простое одностадийное кластерное.

— Расчеты эффекта и его вариации носят более сложный характер.

Многофазный подход к формированию выборки (*multi-phase sampling*). Выборка, формирующаяся в несколько фаз, ограничена единицами и теми же рамками, а единицы исследования имеют одну и ту же структуру в каждой фазе. Собирается базисная информация о единицах наблюдений в большой выборке, а затем для части этих единиц наблюдения (подвыборки) собирается дополнительная информация (рис. 13). Чаще всего предполагается наличие двух фаз или двойное формирование выборки, хотя может быть 3 фазы и больше.

Многофазный подход оказывается полезным в тех случаях, когда рамки популяции не содержат в себе дополнительной информации, необходимой для стратификации или для детального обследования части популяции. Предположим, поставлена задача изучить психологические характеристики родителей, воспитывающих детей первого года жизни, проживающих совместно с бабушками и дедушками, но в распоряжении исследователя имеются списочные рамки, ограниченные информацией о семьях, в которых имеются дети такого возраста. В подобных условиях может быть проведено простое кросс-секционное исследование с единственным вопросом: проживают ли родители и ребенок совместно с бабушками и дедушками? После того как в результате реализации этой первой фазы сформируется выборка из семей, проживающих совместно с бабушками и дедушками, можно будет сформировать из них меньшую по размеру выборку, в которой уже будет проведено углубленное исследование.

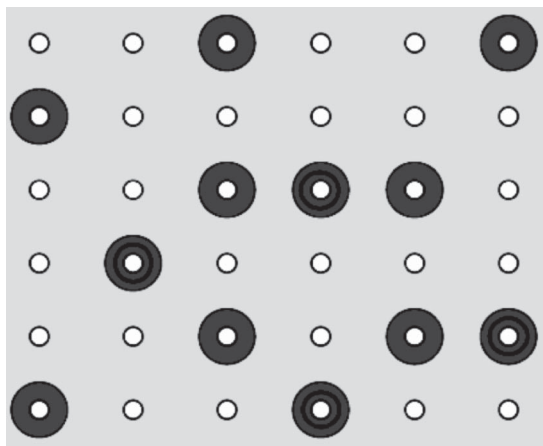


Рис. 13. Многофазный подход к формированию выборки. В ходе первой фазы отобрано 12 единиц наблюдения, в ходе второй — 4 единицы (Franklin S., Walker C., 2010)

Многофазный подход может быть реализован тогда, когда затраты, связанные с получением информации на отдельных этапах исследования, сильно различаются. Например, исследование может быть направлено на изучение важнейших показателей, отражающих пищевое поведение и диетические предпочтения, курение, физическую активность, употребление алкоголя определенного контингента лиц. Кроме того, целесообразно проведение функциональных тестов (определение уровня артериального давления, проведение проб с физической нагрузкой), изучение биохимических показателей (уровня глюкозы и холестерина в сыворотке крови) и т. п. у части таких обследуемых лиц. Очевидно, что простое интервьюирование, направленное на изучение образа жизни, провести гораздо легче, чем реализовать ту часть исследования, которая предполагает использование медицинских манипуляций. В целом исследование может быть проведено в две фазы: базисные вопросы будут заданы испытуемым в первой фазе исследования, углубленное обследование охватит лишь небольшую выборку, сформированную в ходе второй фазы.

Достоинства многофазного подхода к формированию выборки:

- Метод позволяет существенно повысить точность оценок (по сравнению с формированием простой случайной выборки).

- Метод пригоден для получения дополнительной, исходно отсутствовавшей информации, касающейся популяционных рамок.

— Метод используется в тех случаях, когда затраты на получение определенной информации весьма высоки или когда получение этой информации создает большие неудобства для респондента (Franklin S., Walker C., 2010).

Недостатки многофазного подхода к формированию выборки:

— Реализация требует больше времени, чем для осуществления однофазного подхода, если результаты первой фазы необходимы для осуществления второй.

— Метод будет более затратным по сравнению с однофазным подходом, так как требуется повторное обследование участников исследования.

— Если популяция мобильна или интересующие исследователя характеристики часто и быстро меняются, то временные интервалы, возникающие между отдельными фазами исследования, могут создавать трудности для фиксации однотипных и сопоставимых показателей.

— Практическая организация исследования может быть осложнена.

— Расчеты результатов исследования могут быть достаточно сложными.

Реплицированное формирование выборки (*replicated sampling*) предполагает отбор не одной, а нескольких независимых выборок из популяции. Вместо одной последовательно создаются несколько репликатов — выборок меньшего размера, приблизительно равные по численности. Последние формируются независимо друг от друга на основе единого принципа (рис. 14). Реплицированное формирование выборки может использоваться в ситуациях, когда в короткие сроки требуются предварительные результаты на основе обследования и анализа одного репликата. Кроме того, реплицированное формирование выборки позволяет упростить расчет некоторых статистических показателей, прежде всего вариации признака, так как мерой такой вариации будут различия полученных показателей в отдельных репликатах.

В рамках исследований популяционных срезов возможны и специфические варианты дизайна.

Повторные исследования (*repeated surveys*) проводятся с целью изучить тренд или изменения интересующих исследователя характеристик с течением времени. *Тренд* — систематическая направленность в сторону повышения или понижения частоты изучаемого феномена с течением времени. Следует помнить о возможности снижения статистической силы исследования с течением времени. Например, исходно исследование проводилось по принципу стратифицированного формирования выборки со стратификацией по уровню дохода семьи,

однако с течением времени доход мог существенно измениться, и исходные стратификационные рамки, таким образом, могут оказаться нарушенными. Следовательно, при организации повторных исследований такая возможность должна быть предусмотрена.

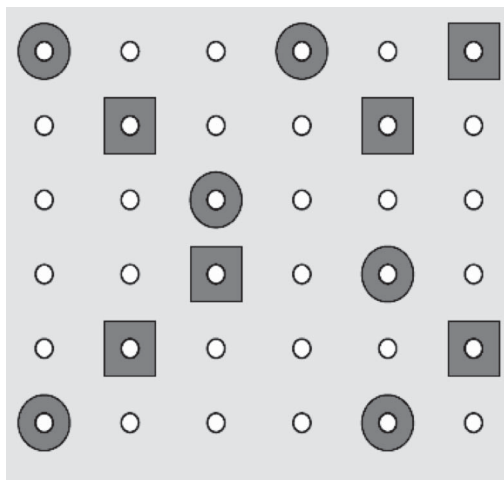


Рис. 14. Схема реплицированного формирования выборки. Представлены две выборки, каждая численностью 6 единиц (Franklin S., Walker C., 2010)

Панельные или лонгитюдные исследования популяционного среза (*panel, longitudinal survey*) — изучение одних и тех же единиц наблюдения в различные моменты времени. Такие исследования обычно направлены на изучение изменения в динамике интересующих исследователя показателей в заданной популяции и позволяют гораздо эффективнее справиться с этой задачей, чем при изучении серии независимых выборок аналогичной численности.

Панельное исследование популяционного среза может также использоваться для оценки надежности психодиагностического метода, поскольку одним из критериев надежности является сходство результатов, полученных при повторных тестированиях (см. главу 3).

В панельном исследовании полезна фиксация относительно редко встречающихся событий.

Пример

В исследовании, посвященном изучению психологических последствий насилия, участники проходили первичное анкетирование, и в ходе анкетиро-

вания изучались признаки, отражающие нынешнее психологическое состояние испытуемых, а также задавался вопрос об имевшемся ранее насилии в отношении испытуемых. В дальнейшем участники проходили анкетирование каждую неделю, причем им задавали вопрос: проявлялось ли насилие по отношению к ним? Если таковое имелося, участники получали дополнительные опросники, направленные на изучение их психологических характеристик, причем эти опросники также предлагалось заполнять повторно: через неделю, две и т. п., что позволяло отслеживать как ближайшие, так и отдаленные последствия насилия (Leeuw E. D. [et al.], 2008).

Панельные исследования популяционных срезов по своему смыслу приближаются к когортным аналитическим исследованиям при определенной модификации. Если в качестве исходно учитываемого признака рассматривается наличие или отсутствие какого-то воздействия, то при динамическом наблюдении можно проследить, ассоциировалось ли данное воздействие с риском возникновения того или иного последствия (эффекта).

Для панельных исследований важнейшей проблемой является отказ респондентов от участия в исследовании, с течением времени таких лиц станет больше, в результате чего численность выборки существенно уменьшается. Также с течением времени может меняться и структура популяции, в результате чего могут добавляться новые единицы наблюдения и выбывать старые. Наконец, панельные исследования труднее реализовать технически, они требуют больших затрат по сравнению с одномоментными исследованиями популяционных срезов.

Формирование выборки по принципу «снежного кома» (*snowball sampling*) является весьма специфическим подходом к исследованию популяционных срезов. Его целесообразность несомненна, когда необходимо обследование лиц с каким-то редким состоянием (заболеванием) или когда единицы наблюдения, которых хотелось бы включить в исследования, не охвачены имеющимися списочными популяционными рамками (например, лица, пишущие стихи, увлекающиеся йогой и т. п.). Исследователь может знать о наличии таких пациентов и имеет возможность связаться с ними и обследовать их. Допустимо спросить этих участников исследования, знают ли они еще кого-нибудь, кто имеет то же заболевание (состояние и т. п.), и, если они таковых знают, можно попросить связаться с ними и предложить им принять участие в исследовании. Таким образом, выборка увеличивается как снежный ком. Принцип «снежного кома» — это не вероятностный подход к формированию выборки, так как не все потенциальные участники исследования имеют равные шансы отбора для участия в нем.

2.1.3. Корреляционные исследования

Цель корреляционных исследований состоит в изучении связи между двумя переменными или более. У каждого испытуемого изучается несколько параметров, чтобы в дальнейшем выявить наличие ассоциации между этими параметрами. Корреляционный дизайн не обязательно предполагает, что в качестве меры ассоциации изучаемых признаков будет выступать коэффициент корреляции, хотя чаще всего так и происходит. Коэффициент корреляции как мера ассоциации имеет недостаток — позволяет изучить лишь один тип ассоциации между переменными — линейную связь. Однако связь между рассматриваемыми переменными может иметь и иной характер, при этом с методологической точки зрения попытка выявить такую связь также будет относиться к корреляционному исследованию. Вот почему некоторые исследователи предлагают подобные дизайны исследований обозначать не как корреляционные, а как *реляционные* (*relational designs*) (Aveline M., Shapiro D., 1995).

Пример

Выявлялась ассоциация между расстройствами сна и выраженностью эмоциональной реактивности у детей 5 лет (Kelmanson I. A., 2013a). На первом этапе было проведено кросс-секционное обследование 150 детей, посещавших детские сады, отобранных по принципу многостадийного (2 стадии) формирования выборки. Провели анкетирование детей с помощью стандартизованного опросника с целью выявить отдельные варианты расстройств сна. Одновременно проводилось анкетирование воспитателей детских садов при помощи опросника Ахенбаха, предназначенного для выявления симптомов эмоциональных нарушений и отклонений поведения детей. Исследование выявило:

- статистически достоверную положительную корреляцию между уровнем эмоциональной реактивности ребенка и общей выраженностью нарушений сна ($r = 0,20$; $P = 0,016$);
- ассоциацию между отдельными симптомами нарушений сна и выраженностью эмоциональной реактивности ребенка (рис. 15).

Очевидно, корреляционное исследование может стать естественным продолжением исследования популяционных срезов, когда задачи исследования не ограничиваются изучением распространенности тех или иных признаков, а предполагают их совокупный анализ. Изучаемые параметры анализируются в один и тот же момент времени. Также корреляционный дизайн может стать естественным продолжением панельного (лонгитюдного) исследования популяционного среза, когда изучение параметров проводится на двух различных временных этапах и более.

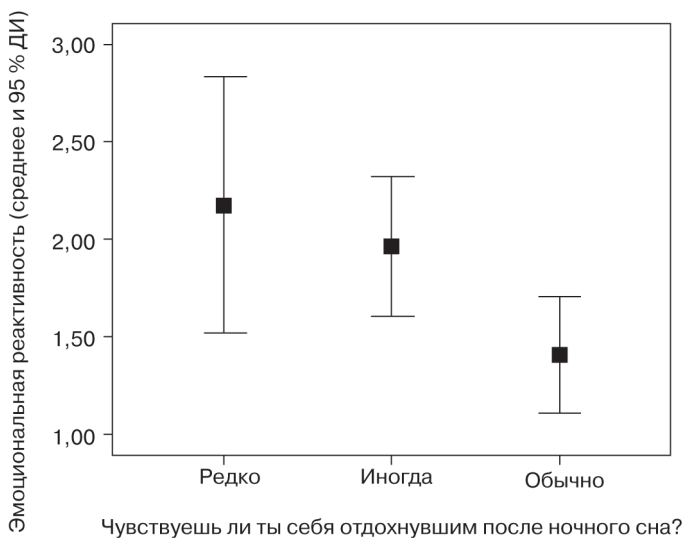


Рис. 15. Связь выраженности эмоциональной реактивности детей 5 лет, исследованной по одной из шкал опросника Ахенбаха, с субъективным ощущением себя отдохнувшим после ночного сна. 95 % ДИ = границы 95 %-го доверительного интервала) (Kelmanson I. A., 2013a)

Мерой ассоциации между показателем распространенности заболевания (патологического состояния или иного изучаемого феномена) и исследуемой характеристикой в корреляционных исследованиях чаще всего выступает коэффициент корреляции. В случаях линейной зависимости параметров используется коэффициент корреляции Пирсона r . Данный показатель отражает выраженность линейной связи между объясняющей переменной и зависимой переменной — изучаемым заболеванием, т. е. говорит о том, в какой мере изменение объясняющей переменной на 1 единицу приводит к пропорциональному увеличению или уменьшению частоты заболевания (патологического состояния). Корреляция может быть положительной или отрицательной, а значение коэффициента корреляций варьирует в диапазоне от -1 до $+1$. Когда связь между независимой и зависимой переменными носит нелинейный характер, в качестве меры ассоциации может использоваться коэффициент ранговой корреляции Спирмена ρ («ро») или Кендалла τ («тау»). Подробное описание методики расчета этих показателей приводится в специальных руководствах

(Howell D. C., 2013; Peat J. K., Barton V., 2014). Корреляционные исследования могут использовать и иные статистические меры выявления ассоциации, например критерий χ^2 . Допустимы многофакторные методы анализа, например множественная регрессия, факторный анализ.

Экологические, или агрегированные, корреляционные исследования сфокусированы не на изучении отдельных индивидуумов, а на обследовании групп в целом. Их предпосылкой является отсутствие данных о совместном распределении минимум двух, а иногда и всех показателей на уровне индивидуальных субъектов. В ходе экологического исследования параметры измеряются на уровне отдельных групп, организаций или географических территорий.

Показатели, используемые в экологических корреляционных исследованиях:

— агрегированные показатели (*aggregate measures*) для какой-то группы индивидуумов (средние значения или пропорции), например распространенность курящих в популяции, средний доход семей и т. п.;

— показатели, отражающие характеристики окружения (*environmental measures*), т. е. некоторые характеристики того места, где находятся представители изучаемой группы (например, загрязненность воздуха, число солнечных дней в году и т. п.). Каждый такой показатель может быть изучен у конкретного индивидуума, и поэтому индивидуальная подверженность испытываемому такому фактору может варьировать;

— показатели, отражающие глобальные характеристики (*global measures*), являющиеся атрибутами групп, организаций, отдельных территорий, для которых отсутствуют аналоги на индивидуальном уровне (например, плотность населения, рождаемость, смертность, тип системы здравоохранения и т. п.) (Rothman K. J. [et al.], 2008).

Уровни анализа, в экологических корреляционных исследованиях:

— *Индивидуальный уровень анализа (individual-level analysis)*. Величина каждой переменной присваивается каждому отдельному испытуемому. Например, средний уровень загрязненности воздуха в соответствующем регионе может «присваиваться» каждому участнику исследования, проживающему на данной территории, и рассматриваться в качестве возможного предиктора развития патологии.

— *Полностью экологический уровень анализа (completely ecologic analysis)*. Все переменные (предполагаемое воздействие, его эффект и дополнительные влияющие факторы) рассматриваются как экологические показатели, единицей анализа выступает не отдельный индивид, а группа (например, страна, регион, школа, медицинское учреждение, демографическая страта или временной промежуток). Таким

образом, распределение показателей у каждого отдельного индивида не учитывается. Параметры, характеризующие популяцию в целом, используются для описания распространенности заболевания (патологического состояния, исследуемого феномена).

Пример

Примером может служить изучение заболеваемости в зависимости от года, сезона, дней недели, возраста обследуемых, характеристик уровня медицинской помощи, особенностей пищевого рациона, распространенности отдельных вредных привычек и т. п. Так, было проведено исследование распространенности курения среди подростков в США и ее сопоставление с частотой демонстрации эпизодов курения в 25 наиболее кассовых фильмах, находящихся в кинопрокате (Sargent J. D., Heatherton T. F., 2009). Выявлено, что между частотой демонстрации эпизодов курения в кинофильмах и распространенностью курения среди подростков существуют зависимость, и по мере снижения демонстрации курения в кинофильмах отмечается распространенность этой вредной привычки среди подростков. Подобные наблюдения позволили высказать предположение, что демонстрация табачной продукции в кинофильмах является барометром социального одобрения курения, и необходимо направить усилия на ограничение такого воздействия на подростков (Sargent J. D., Heatherton T. F., 2009) (рис. 16).

Достоинства корреляционных исследований:

- Исследования достаточно легко осуществимы.
- Необходимая информация, как правило, уже имеется в наличии и доступна. Однако такого рода исследования имеют и ряд недостатков.

Недостатки корреляционных исследований:

- Корреляционные исследования не позволяют проанализировать связь между предполагаемым патогенным воздействием и развитием заболевания (патологического состояния) у конкретного индивидуума.

— Корреляционные исследования часто не позволяют учесть влияние иных факторов, которые также могут быть связаны с риском возникновения заболевания.

— Выявленная корреляция не означает непосредственной причинно-следственной (каузальной) связи между рассматриваемым заболеванием и изучаемым показателем, выступающим в качестве независимой переменной.

— Данные, полученные в экологических корреляционных исследованиях, представляют сведения о среднем уровне того или иного патогенного воздействия в популяции, а не о реальном уровне данного воздействия на каждого отдельно взятого обследуемого. Поэтому наличие в целом положительной или отрицательной корреляции (ас-

социации), выявляемой по данным корреляционных исследований, еще не позволяет судить о возможном существовании более сложной взаимосвязи между уровнем патогенного воздействия и риском развития заболевания.

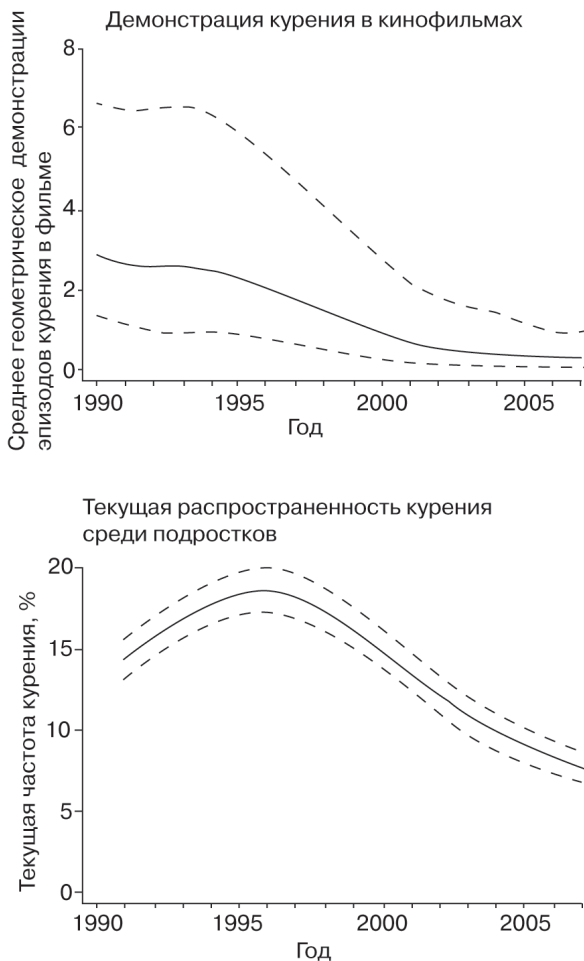


Рис. 16. Связь числа эпизодов демонстрации курения в 25 наиболее «кассовых» фильмах в США и текущей (за последние 30 дней) распространенности курения среди подростков. Пунктирные линии соответствуют 95 %-му ДИ (Sargent J. D., Heatherton T. F., 2009)

Пример

Изучалась связь распространенности курения среди мужчин со стоимостью табачной продукции. Проанализировав данные Всемирной организации здравоохранения о том и другом показателе в различных регионах мира, автор исследования выявил наличие статистически достоверной отрицательной корреляции: по мере роста стоимости пачки сигарет снижался процент курящих мужчин (Townsend J., 2015) (рис. 17). Выявленную ассоциацию можно объяснить тем, что рост стоимости табачной продукции снижает ее доступность и наоборот. И таким образом отражается на распространенности в популяции этой вредной привычки. Однако нет оснований с уверенностью утверждать, что низкая стоимость табачной продукции является непосредственной причиной высокой частоты курения в популяции: возможно, низкая стоимость и доступность табачной продукции являются маркером присутствия иных, ассоциированных с ней и неучтенных факторов, которые, в свою очередь, действительно влияют на распространенность курения в популяции. В частности, доступность табачной продукции может в целом отражать более низкий социально-экономический уровень региона, негативно отражающийся на психологическом состоянии населения, что провоцирует аддиктивное поведение. Кроме того, относительная доступность табачной продукции может свидетельствовать о более низком уровне санитарного просвещения, что также способствует росту распространенности курения. Возможны и иные объяснения. Ответить на вопрос, какое из предложенных объяснений наиболее правдоподобно, не представляется возможным, опираться только на данные корреляционного исследования.

Для попытки ответа на подобные вопросы возможно применение более сложных подходов к анализу данных, в частности **каузальное моделирование, или структурное моделирование уравнений** (Kline R. B., 2011). Смысл этого подхода состоит в оценке того, насколько хорошо модель, предполагающая определенные причинно-следственные связи, разработанная на основе имеющихся предшествующих результатов исследований или теоретических предположений, соответствует фактическим данным. Структурное моделирование является инструментом для осуществления попытки концептуально осмыслить результаты корреляционного анализа.

В ходе проведения экологических корреляционных исследований элементом анализа часто является изучение наличия или отсутствия тренда и его отграничения от флюктуаций (случайных колебаний частоты). Наличие статистически достоверного тренда может быть связано с влиянием ряда факторов, к важнейшим из которых можно отнести:

— улучшение качества диагностики, что приводит к увеличению числа распознаваемых случаев заболевания даже при отсутствии истинного возрастания заболеваемости;

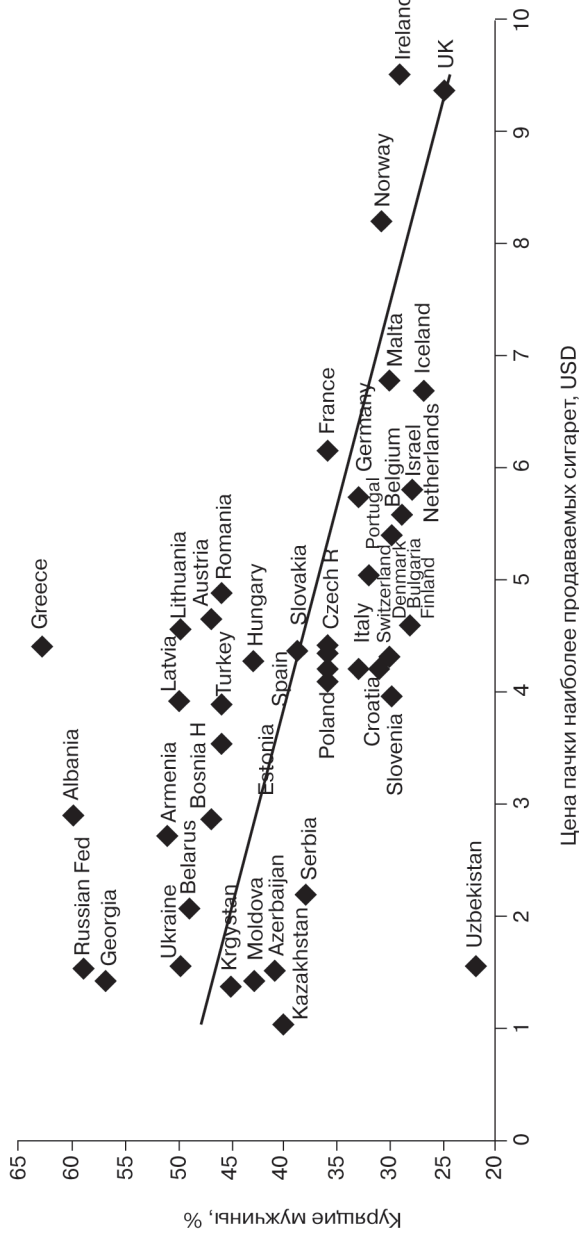


Рис. 17. Связь распространенности курения среди мужчин со стоимостью табачной продукции в различных странах (данные ВОЗ) (Townsend J., 2015)

— появление более строгих подходов к формированию популяции, для которой характерен риск развития данного заболевания, что приводит к изменению показателей, хотя также не означает истинного изменения частоты заболевания;

— изменения возрастного состава населения, что может приводить к изменениям показателей даже в тех случаях, когда частота, специфичная для отдельных возрастных категорий, остается прежней;

— изменения выживаемости пациентов, связанные с улучшением качества лечения или более ранней диагностикой;

— истинные изменения частоты новых случаев заболевания, что может быть связано с изменениями внешних факторов и образа жизни (Кельмансон И. А., 2002).

Чаще всего оценка статистической достоверности наличия или отсутствия линейного тренда основана на использовании корреляционного и регрессионного анализа: уравнения линейной регрессии, где в качестве независимой переменной выступает год наблюдения, или критерия Манна — Кендалла, в основе которого лежит непараметрический критерий корреляции Кендалла (Hollander M. [et al.], 2014).

Пример

Данные о динамике зарегистрированных случаев суицида в США с 1986 по 2000 г. (Американский фонд по профилактике суицида, www.afsp.org/understanding-suicide/facts-and-figures) (рис. 18) показывают тенденцию к снижению частоты случаев суицида: со 130 до 104 на 1 млн населения. Расчеты позволяют говорить о наличии статистически достоверного тренда, причем в уравнении линейной регрессии год выступает в качестве независимой переменной, а регистрируемая частота — в качестве зависимой: частота = $3542,16 - 1,72 \times (\text{год})$. Коэффициенту линейной регрессии ($-1,72$) соответствует значение $P < 0,0001$ (высокий уровень статистической достоверности). Сходным образом использование критерия Манна — Кендалла также свидетельствует о наличии статистически достоверного тренда ($\tau = -0,961$; $P < 0,002$).

Большая часть статистических программ позволяет выполнить необходимые расчеты. Имеются и специализированные бесплатные компьютерные программы, ориентированные на эпидемиологический анализ, позволяющие выполнять указанные расчеты. В частности, можно отметить программы WINPEPI (<http://www.brixtonhealth.com>) (Abramson J. H., 2011) и Joinpoint (<http://surveillance.cancer.gov/joinpoint>) (Statistical Research and Applications Branch, National Cancer Institute, 2015).

Joinpoint выявляет наличие достоверного тренда и анализирует наличие возможных «точек излома» (*join points*), отражающих изменение тренда (Kim H. J. [et al.], 2000).

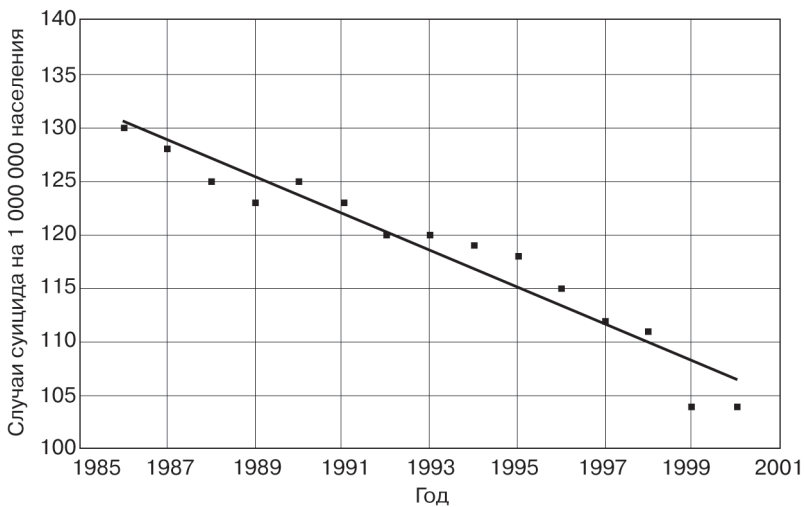


Рис. 18. Динамика случаев суицида в США с 1986 по 2000 г. (www.afsp.org/understanding-suicide/facts-and-figures)

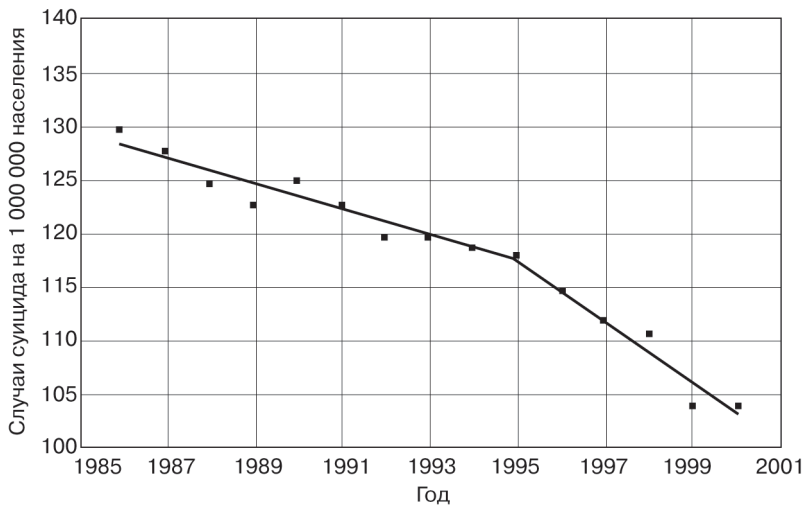


Рис. 19. Динамика случаев суицида в США с 1986 по 2000 г. (www.afsp.org/understanding-suicide/facts-and-figures). Модель с наличием «точек излома»

Пример

Анализ статистических данных, отражающих динамику случаев суицида в США за 1986–2000 гг., позволил выявить наличие одной «точки излома», учет которой позволяет оптимальным образом описать динамику показателей (рис. 19). Данная модель позволяет выявить наличие двух сегментов:

– первый (1986–1995 гг.) описывается уравнением регрессии:

Число случаев суицида на 1 млн = $2564,33 - 1,23 \times (\text{год})$,

где статистическая достоверность коэффициента регрессии $P = 0,000046$;

– второй сегмент (1995–2000 гг.) характеризуется еще более крутым снижением частоты случаев суицида:

Число случаев суицида на 1 млн = $5793,97 - 2,85 \times (\text{год})$,

где статистическая достоверность коэффициента регрессии $P = 0,000073$.

Пример

Анализируется динамика рождаемости в России с 1990 по 2008 г. (данные Федеральной службы государственной статистики: http://www.gks.ru/scripts/db_inet/dbinet.cgi?pl=2401002) (рис. 20). В целом можно отметить снижение рождаемости, однако детальный анализ позволяет выявить две точки излома, которые оптимальным образом характеризуют динамику показателей:

– первый сегмент (1990–1993 гг.) показывает резкое снижение рождаемости с наличием статистически достоверного тренда, показатели рождаемости описываются уравнением линейной регрессии:

Число родившихся живыми = $389350184,27 - 194654,86 \times (\text{год})$;

$P = 0,000059$;

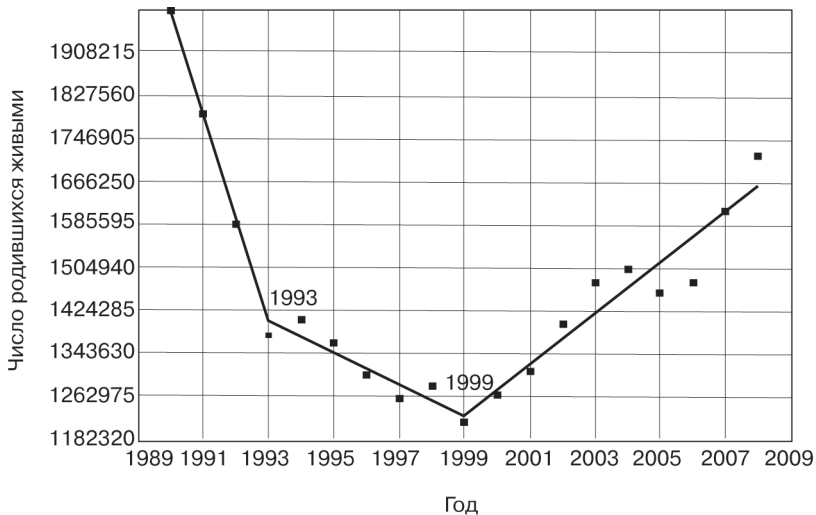


Рис. 20. Рождаемость в России в 1990–2008 гг. По данным Федеральной службы государственной статистики: http://www.gks.ru/scripts/db_inet/dbinet.cgi?pl=2401002

– второй сегмент (1993–1999 гг.) отмечен снижением рождаемости, но не столь резким, как на первом этапе, и уравнение регрессии выглядит следующим образом:

$$\begin{aligned} \text{Число родившихся живыми} &= 59072541,40 - 28936,02 \times (\text{год}); \\ P &= 0,061. \end{aligned}$$

Как видно из значения P , тренд на этом сегменте не достигает уровня статистической достоверности.

– третий сегмент (1999–2008 гг.) позволяет констатировать статистически достоверный тренд в сторону роста рождаемости:

$$\begin{aligned} \text{Число родившихся живыми} &= -94064460,12 + 47670,78 \times (\text{год}); \\ P &= 0,000004. \end{aligned}$$

Нередко корреляционные исследования используются для выявления сезонности динамики тех или иных показателей, что также может иметь большое значение для понимания изучаемых процессов. Большинство статистических программ позволяет выполнить подобные расчеты, в частности упомянутая выше программа WINPEPI.

Для выявления статистической достоверности различий показателей T , наблюдаемых в отдельные месяцы, может быть использован критерий χ^2 (хи-квадрат), сопоставляющий фактическое распределение с теоретически ожидаемым. Рассчитываемый показатель будет выглядеть следующим образом:

$$T = \sum_{i=1}^k \left[\frac{(O_i - E_i)^2}{E_i} \right],$$

где O_i – фактическое значение наблюдений в i -й месяц; E_i – ожидаемое значение в i -й месяц; k – число месяцев ($k = 12$). Теоретически ожидаемое значение E_i рассчитывается исходя из нулевой гипотезы, т. е. из предположения о равномерном распределении показателей в течение года (ежемесячное значение составляет: $100\% / 12 = 8,3\%$, если не принимать во внимание различную продолжительность отдельных месяцев).

Сходные оценки дает использование критерия Фридмана (*Freedman's test*), близкого по смыслу к критерию Колмогорова – Смирнова.

Достоверность наличия периодических колебаний показателей (с периодом 4, 5 и 6 мес.) может быть оценена при помощи критериев Эдвардса и Хьюитта (*Edwards's test*, *Hewitt test*). Критерий Хьюитта используется, когда анализируется ранговое распределение месячных показателей. Месяцу с максимальным числом наблюдений присваивается ранг, равный 12, месяцу с наименьшим показателем – ранг, равный 1. Для каждого анализируемого периода времени (4, 5

и 6 мес.) рассчитываются все возможные суммы рангов месяцев, следующих друг за другом. Так, для анализируемой шестимесячной периодичности анализируется сумма рангов месяцев с января по июнь, с февраля по июль, с марта по август и т. д. Группа из 6 мес. с максимальной полученной суммой рангов рассматривается как пик, а статистическая достоверность оценивается на основании величины этой суммы рангов. Подробно методика расчета этих критериев изложена в специальных руководствах (Helsel D. R., Hirsch R. M., 1992; Rau R., 2007).

Пример

Анализ данных о помесечном распределении зарегистрированных случаев суицида в Чили с 1979 по 1994 г. (Retamal C. P., Humphreys D., 1998) (рис. 21) свидетельствует об отсутствии статистически достоверного отклонения помесечных показателей от теоретически ожидаемого равномерного распределения ($P > 0,1$ по критерию Фридмана). В то же время можно говорить о наличии четырехмесячного пика с октября по январь ($P = 0,024$ по критерию Хьюитта), пятимесячного пика с сентября по январь ($P = 0,015$) и шестимесячного пика с сентября по февраль ($P = 0,013$).

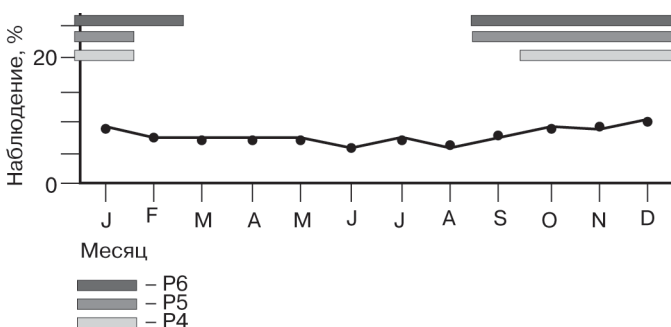


Рис. 21. Сезонность случаев суицида в Чили в 1979—1994 гг.: P6, P5, P4 — периодичность, 6, 5 и 4 мес. соответственно (Retamal C. P., Humphreys D., 1998)

2.2. Аналитические исследования

Аналитический дизайн фокусирует свое внимание на факторах, предположительно определяющих риск развития заболевания, и в идеальном случае позволяет дать ответ на вопрос, ассоциировано ли конкретное воздействие (*exposure*) с эффектом (*effect*), реализующимся в виде увеличения или уменьшения риска развития того или

иного изучаемого заболевания (состояния). Следует иметь в виду, что термин «воздействие» может трактоваться весьма широко:

- присутствие какого-то физического или химического фактора;
- наличие любого иного признака, который потенциально может повлиять на риск возникновения заболевания (например, масса ребенка при рождении, группа крови и т. п.). Исследователь целенаправленно формирует группы обследуемых индивидов для систематического анализа и поиска ответа на вопрос: различается ли риск возникновения заболевания в зависимости от того, подвергались или не подвергались обследуемые влиянию интересующего воздействия? Для аналитического исследования критичным является формирование адекватной контрольной группы наблюдений, что позволяет проверить достоверность рабочей гипотезы.

Имеется широкий выбор подходов к планированию аналитических исследований, однако принципиально следует выделить два варианта, различающиеся по роли исследователя в ходе их проведения:

- *Обсервационные исследования.* Исследователь только наблюдает за естественным развитием событий, беспристрастно фиксирует информацию о том, кто из обследованных лиц подвергался изучаемому воздействию, а кто нет; у кого отмечался эффект такого воздействия, а у кого — нет.

- *Интервенционные исследования.* Исследователь сам определяет, кто из обследуемых лиц должен быть подвержен изучаемому воздействию, а кто — нет. В дальнейшем он наблюдает за обследуемыми лицами, констатирует возникновение или отсутствие у них интересующего его исхода (Кельмансон И. А., 2002; Rothman K. J. [et al.], 2008).

2.2.1. Обсервационные аналитические исследования

2.2.1.1. Исследования по принципу «случай — контроль»

Исследования по принципу «случай — контроль» (*case-control studies*) относятся к категории обсервационных аналитических исследований, в ходе проведения которых обследуемые субъекты отбираются в зависимости от наличия («случай») или отсутствия («контроль») у них анализируемого заболевания (состояния). Формируются соответствующие группы — основная и контрольная, где сопоставляется частота встречаемости предшествовавшего предполагаемого воздействия, чаще всего патогенного характера (Кельмансон И. А., 2002; Покровский В. П., Брико Н. И., 2012). Основную группу можно также именовать по профилю патологии пациентов (например, дети с нарушениями сна). Иногда говорят о дисфункциональной (случай) и функциональной (контроль) группах (Barker C. [et al.], 2002).

Достоинства исследований по принципу «слугай — контроль»:

— При изучении заболеваний с длительным латентным периодом значительно легче выявить лиц со сформировавшимся заболеванием, у которых возможно ретроспективное изучение интересующих анамнестических данных, нежели «ожидать» в течение многих лет развития заболевания среди лиц, исходно вошедших в анализируемую когорту.

— Формирование изучаемых групп в зависимости от статуса по отношению к наличию заболевания позволяет исследователю произвольно регулировать численность групп. Следовательно, подобные исследования весьма эффективны в случае анализа редко встречающихся заболеваний (состояний).

— Возможен анализ большого числа потенциальных предрасполагающих факторов риска и изучение их взаимодействия. Исследования могут быть использованы для проверки гипотезы о наличии определенных факторов риска. При отсутствии такой предварительной гипотезы исследования по принципу «случай — контроль» позволяют дать оценку выраженности различий частоты, с которой те или иные потенциально патогенные воздействия наблюдались среди лиц, имеющих конкретное заболевание и не имеющих такового. Такое сопоставление важно при изучении патологических состояний с окончательно неустановленной этиологией (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008).

Недостатки исследований по принципу «слугай — контроль»:

— К моменту формирования сопоставляемых групп уже есть как анализируемые воздействия, так и развитие заболевания. Возможно влияние неравноценного подхода к отбору «случаев» и контрольных наблюдений в зависимости от имевшегося патогенного воздействия, различного представления информации об имевшемся в анамнезе патогенном воздействии в зависимости от наличия или отсутствия анализируемого заболевания. Таким образом, в таких исследованиях предъявляются самые жесткие требования к источникам информации, и от надежности этих источников во многом зависит то, в какой мере удастся избежать ошибок.

— Является проблемой установление последовательности событий во времени.

— Проблематично определить, в какой мере «случаи», ставшие объектом исследования, репрезентативны по отношению ко всем лицам в популяции, имеющим данное патологическое состояние, а полученные результаты обобщаемы. Подобные исследования могут быть ограничены лишь отдельными формами заболевания, в отношении которых представляется возможным получить полную и достоверную информацию об имевшемся воздействии и возникшем заболе-

вании. Если лица, отобранные для исследования в качестве «случаев», существенно отличаются от иных пациентов по ряду показателей (например, по возрасту, полу, этнической принадлежности, степени тяжести заболевания), то весьма вероятно, что и выраженность предшествовавшего патогенного воздействия в этих наблюдениях также будет существенно отличаться от фактически имевшей место у всех лиц, страдающих этим заболеванием.

Принципиальным моментом при проведении исследования по принципу «случай — контроль» является существование критериев для суждения о наличии или отсутствии заболевания (патологического состояния). Важно, чтобы анализируемое заболевание (состояние) было представлено максимально гомогенно, так как нередко различные по своей природе заболевания имеют сходные клинические проявления. В целях формирования гомогенной группы наблюдений первым этапом должна быть формулировка строгих диагностических критериев. Идеально, если исследователь может опираться на так называемый «золотой стандарт» диагностики. Нередко используются принятые международные критерии диагноза, например критерии Американской психологической ассоциации. В силу диагностических трудностей возможны ситуации, когда невозможно с абсолютной уверенностью говорить о наличии того или иного диагноза, тогда полезен отдельный анализ «случаев» с различной степенью вероятности диагноза. После установления диагностических критериев становится возможным отбор наблюдений из различных источников:

— *Исследование по принципу «случай — контроль», основанное на госпитальных данных (hospital-based case-control study)*, проводится среди лиц, госпитализировавшихся в конкретное медицинское учреждение с тем или иным диагнозом в течение фиксированного периода времени.

— *Исследование по принципу «случай — контроль», основанное на популяционных данных (population-based case-control study)*, требует либо информации о всех тех лицах, которые страдают данным заболеванием в определенной популяции, либо сведений, касающихся случайной выборки (random sample) таких случаев. Например, это могут быть лица, страдающие данным заболеванием, которые являются представителями популяции в целом на тот или иной момент времени.

Достоинства данного исследования таковы:

— возможность избежать влияния субъективизма, связанного с принципом отбора обследуемых;

— возможность воспроизвести полную картину распространенности заболевания в данной популяции и непосредственного подсчета частоты заболевания среди лиц, подвергшихся и не подвергшихся анализируемому воздействию.

Инцидентные исследования по принципу «слугай — контроль» (incident case-control studies). Обследуются лица, имеющие анализируемое заболевание, у которых впервые установлен диагноз (**инцидентные случаи**). При планировании исследований по принципу «случай — контроль» следует отдавать предпочтение анализу новых (инцидентных) случаев заболевания, диагностированных в течение ограниченного периода времени.

Превалентные исследования по принципу «слугай — контроль» (prevalent case-control studies) (Morgenstern H., Thomas D., 1993). Обследуются лица, уже имеющие заболевание на определенный момент времени (**превалентные случаи**).

При изучении превалентных случаев возможно существенное увеличение численности выборки, доступной для анализа в данный момент времени, особенно при редко встречающихся заболеваниях. Данное исследование может быть составной частью исследования популяционного среза, причем последнее может предполагать дифференцированный отбор обследуемых в зависимости от наличия или отсутствия у них изучаемого заболевания с тем, чтобы увеличить число лиц с наличием заболевания в выборке. Однако такие превалентные наблюдения будут подвержены воздействию факторов, влияющих как на возникновение новых случаев заболевания, так и на продолжительность заболевания. В подобной ситуации ответить на вопрос о том, в какой степени предполагаемое воздействие влияет на риск развития заболевания и в какой на его прогноз, можно лишь тогда, когда точно известен эффект, оказываемый данным воздействием на длительность заболевания и его течение. События, рассматриваемые в качестве потенциальных провоцирующих факторов, должны предшествовать заболеванию, а не быть следствием его течения. Данное условие сложно выполнить, если в исследование включают все зарегистрированные случаи заболевания, а не впервые диагностированные.

Пример

Изучается возможная связь между уровнем употребления кофе и риском развития нарушений сна. Возможно, будет трудно понять, в какой мере сведения об уровне употребления кофе, сообщенные пациентами, страдающими нарушениями сна, отражают ситуацию, предшествовавшую развитию заболевания, а в какой мере отражают их изменившиеся привычки в результате появления начальных признаков заболевания, возможно, еще до точного установления диагноза.

Контрольные наблюдения необходимы для того, чтобы имелась возможность оценить, насколько частота предполагаемого патогенного воздействия у лиц, страдающих изучаемым заболеванием,

отличается от той, которую можно было бы ожидать, исходя из ситуации, наблюдаемой среди лиц с отсутствием признаков этого заболевания. Адекватный подбор контрольных наблюдений является одним из наиболее трудных и во многом критических моментов при проведении исследований по принципу «случай — контроль».

Оптимальная методика формирования контрольной группы наблюдений зависит от конкретной задачи. Отбор контрольных наблюдений предполагает учет характеристик «случаев» и принципа их отбора, необходимости получения сопоставимой информации у «случаев» и контрольных наблюдений, а также от иных чисто практических соображений. В зависимости от того, из каких источников формируется группа случаев (например, дети, наблюдаемые конкретным участковым педиатром; все больные дети района, города и т. п.), будет зависеть и подход к формированию контрольных наблюдений.

Контрольные наблюдения должны представлять не всех членов популяции, у которых отсутствует заболевание, а лишь тех, у которых отсутствует заболевание и которые потенциально могли быть включены в исследование в качестве «случаев», если бы у них имелось это заболевание. Следовательно, все те ограничения, которые накладываются на отбор «случаев», должны распространяться и на контрольные наблюдения, и наоборот.

При формировании контрольной группы наблюдений возможны различные варианты:

- госпитальные контрольные наблюдения;
- контрольные наблюдения, основанные на общепопуляционных данных.

Каждый источник имеет свои определенные преимущества и недостатки, которые следует учитывать в зависимости от задач конкретного исследования, принципа отбора случаев и той информации, которая должна быть проанализирована.

Госпитальные контрольные наблюдения. Если «случаи» отбираются на основе госпитальных данных, нередко в качестве контрольных наблюдений подбираются лица, госпитализированные в те же больничные учреждения, но с иными диагнозами.

Достоинства госпитальных контрольных наблюдений:

— Контрольные наблюдения легко идентифицируемы и доступны для анализа в достаточном количестве.

— Факт госпитализации предполагает, что пациенты с большей точностью могут воспроизвести информацию об имевшихся в прошлом потенциальных патогенных воздействиях, чем здоровые лица, находящиеся вне больничных учреждений. Контрольные наблюдения достаточно сопоставимы с лицами, вошедшими в группу «случаев».

ев», а вероятность ошибки воспроизведения информации в данной ситуации сводится к минимуму.

— Обследуемые могли в той же степени, что и «случаи», подвергнуться влиянию факторов, которые предопределили их обращение в данное медицинское учреждение или к конкретному врачу.

— Обследуемые с большей вероятностью будут склонны к сотрудничеству, нежели здоровые лица, что приведет к минимизации ошибки, связанной с нежеланием участвовать в исследовании.

Недостатки госпитальных контрольных наблюдений:

— Обследуемые лица не являются здоровыми и по ряду параметров, связанных с заболеванием и госпитализацией, могут существенно отличаться от практически здоровых людей.

— Нередко на госпитализацию оказывают сильное влияние социальные факторы.

Возможна специализация стационара на лечении заболеваний определенного профиля, в нем собраны лица, страдающие заболеваниями определенного класса. В подобной ситуации больница является превосходным источником для отбора «случаев», которые, однако, могут быть несопоставимы с иными пациентами, госпитализированными в ту же больницу. Так, «случаи» могут стекаться из различных регионов или быть выходцами из более обеспеченных семей, которые могут себе позволить лечение в специализированном стационаре.

Актуален вопрос о том, наличие каких форм заболеваний можно считать допустимым для включения в группу госпитальных контрольных наблюдений. Ряд заболеваний, по поводу которых такие пациенты госпитализируются, может быть результатом влияния провоцирующих факторов, с которыми ассоциировано и основное изучаемое заболевание. Следовательно, лица с патологией, имеющей доказанную ассоциацию (как положительную, так и отрицательную) с анализируемыми факторами риска, не должны включаться в контрольную группу наблюдений. Так, курение матери во время беременности является доказанным фактором риска развития бронхиальной астмы у ребенка. Если при изучении вероятной ассоциации между курением матери во время беременности и риском формирования дефицита внимания и гиперактивности у ребенка включить в контрольную группу наблюдений детей, страдающих бронхиальной астмой, то пропорция курящих матерей в контрольной группе будет заведомо завышенной, и такая контрольная группа не будет адекватно представлять популяцию в целом. Показатели риска, полученные на основании такого исследования, будут занижены. Кроме того, лица, у которых развилось патологическое состояние могло привести к изменению уровня предполагаемого патогенного воздействия, также должны быть исключены из исследования. Так, при изучении

эффектов (например, нарушений сна), связанных с употреблением кофе, в контрольную группу наблюдений не следует включать лиц, страдающих гастродуоденальной патологией, так как заболевания желудочно-кишечного тракта могут способствовать диетическим ограничениям в рационе, в частности отказу от употребления кофе (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987).

Контрольные наблюдения, основанные на общепопуляционных данных. Если «случаи» были сформированы на основании госпитальных данных, но использование госпитальных контрольных наблюдений нецелесообразно или трудноосуществимо, группа сравнения может быть сформирована на основании общепопуляционных данных. Этим же принципом следует руководствоваться при формировании контрольных наблюдений в тех ситуациях, когда «случаи» отбирались из представителей определенной популяции. Отбор контрольных наблюдений на основе общепопуляционных данных может осуществляться различными методами: обследованием соседей, случайным отбором по номерам телефонов, идентификацией обследуемых по данным медицинских регистров и т. п. Данная задача сопряжена с определенными трудностями:

— Идентификация и опрос контрольных наблюдений, выбранных из общей популяции, всегда требуют больших затрат времени и ресурсов, а список лиц, формирующих популяцию, не всегда доступен.

— Практически здоровые лица, составляющие контрольную группу наблюдений, с большей вероятностью могут упустить какие-то детали своего анамнеза, о которых нередко вспоминают больные.

— Практически здоровые люди менее склонны к участию в научном исследовании, и процент отказов от участия в исследовании среди них существенно выше, чем среди больных («случаев») и госпитальных контрольных наблюдений.

— Практически здоровые лица, соглашающиеся на участие в исследовании, нередко имеют какие-то специфические побудительные мотивы, что также не может не отразиться на полученных результатах (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987).

Контрольные наблюдения с участием специальных групп. К числу таких групп можно отнести друзей, соседей, родственников и т. п. С общепопуляционными контрольными наблюдениями их сближает тот факт, что такие лица, как правило, практически здоровы.

Достоинства контрольных наблюдений с участием специальных групп:

— Специальные группы более склонны к сотрудничеству вследствие своего особого отношения к «случаям».

— Участие специальных групп позволяет во многом нивелировать влияние таких факторов, как этническая принадлежность и генетические особенности, социальные условия, микроокружение.

Недостатки контрольных наблюдений с участием специальных групп:

— Признаки, сближающие обследуемых лиц, являются вероятными факторами риска заболевания.

Количество контрольных групп. После определения оптимального источника формирования контрольных наблюдений требуется решения вопрос о количестве контрольных групп. Необходимо стремиться к формированию единой контрольной группы наблюдений, максимально соответствующей обследуемым «случаям» и таким образом максимально удовлетворяющей задачам исследования. Однако по ряду соображений, особенно при использовании госпитальных контрольных наблюдений, когда нет уверенности в полной независимости статуса обследуемого от вероятных патогенных факторов, возможно формирование нескольких контрольных групп и последующее сопоставление полученных результатов. Очевидно, принципиальное совпадение результатов сопоставлений с различными контрольными группами может явиться аргументом в пользу обоснованности выводов. Различия, выявляемые в результатах сопоставлений с отдельными контрольными группами, могут дать информацию о характере ассоциации и причине вероятной ошибки. Использование нескольких контрольных групп оправдано и в тех ситуациях, когда возникают опасения, что одна из групп может иметь какую-то отличительную особенность. Например, при анализе связи между употреблением кофе и риском нарушений сна контрольная группа наблюдений может формироваться на основе госпитальных данных. Однако при этом высока вероятность, что лица, госпитализированные с иными, отличными от нарушений сна диагнозами, могут в силу имеющихся у них заболеваний исключить из рациона ряд пищевых продуктов, в том числе кофе, и, чтобы избежать подобной ошибки, разумно сформировать вторую контрольную группу наблюдений на основе популяционных данных.

Количество контрольных наблюдений, которые следует включать в анализ. Если численность группы «случаев» достаточно велика, а сбор информации сопряжен с большими трудностями, оптимальным может быть соотношение «случаев» и контрольных наблюдений 1 : 1. Если численность группы «случаев» невелика, затраты, связанные с получением необходимой информации, неодинаковы для «случаев» и контрольных наблюдений, возможно формирова-

ние иной пропорции. По мере увеличения количества контрольных наблюдений, приходящихся на 1 «случай», увеличивается информативность исследования. Если число контрольных наблюдений, приходящихся на 1 «случай», равно r , то процент теоретически доступной информации приблизительно составит $r / (r + 1)$ (Rothman K. J. [et al.], 2008) (рис. 22). Увеличение отношения числа контрольных наблюдений к числу «случаев» более 4 не приводит к существенному увеличению информативности исследования, и поэтому превышение данного отношения мало оправдано.

Формирование случайной выборки заданной численности возможно в тех ситуациях, когда имеются сведения о всей популяции потенциальных контрольных наблюдений. Если группу потенциальных контрольных наблюдений можно организовать по какому-то принципу в виде списка, то возможен отбор каждого десятого, сотого, тысячного объекта из этого списка.

Принцип парных выборок (paired sampling) применяется, если не представляется возможным получить полный списочный состав тех лиц, которые могут выступать в качестве контрольных наблюдений, и если нецелесообразно ожидать момент, когда будет полностью завершен процесс формирования такого списка (например, завершение календарного года). Суть данного метода сводится к тому, что каждому случаю подбирается заранее оговоренное количество соответствующих контрольных наблюдений, причем подбор осуществляется на основании определенных критериев, определяющих временную или

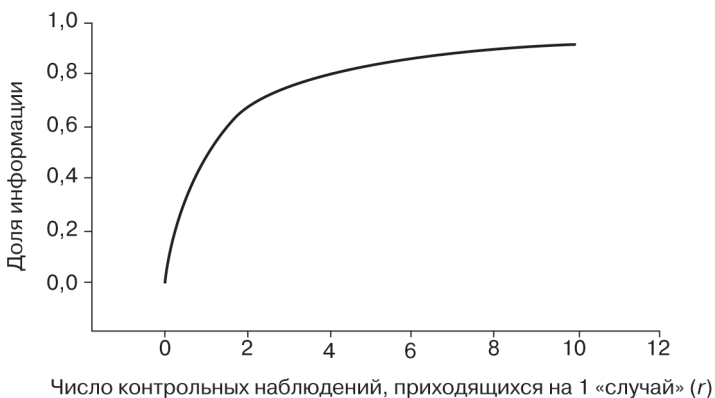


Рис. 22. Зависимость теоретически доступной информации от числа контрольных наблюдений, приходящихся на 1 «случай»

географическую связь со «случаями». Например, при формировании госпитальных контрольных наблюдений может быть использован принцип отбора каждого следующего госпитализируемого пациента того же возраста и пола, что и «случай», или даже проживающего по соседству со «случаем».

Независимо от того, какой принцип используется при отборе контрольных наблюдений, важно, чтобы он был заранее четко сформулирован и последовательно осуществлялся. Непоследовательность в реализации принципа отбора контрольных наблюдений может существенно исказить результаты. В частности, исключение лиц, проявляющих нежелание участвовать в исследовании, может явиться источником систематической ошибки, если причина, с которой связано это нежелание, имеет какое-либо отношение к тому изучаемому патогенному воздействию, которое является предметом анализа в ходе исследования.

Наличие заболевания и анализируемого патогенного воздействия у обследуемых. Соответствующая информация необходима после уточнения характеристик, лежащих в основе отбора «случаев» и контрольных наблюдений, а также после выбора источника формирования обследуемых групп. Ее содержат многочисленные источники: свидетельства о смерти, протоколы патологоанатомических исследований, медицинские регистры, истории болезни, выписные справки, поликлинические карты и т. п. В отношении сбора данных о наличии анализируемого патогенного воздействия также существуют многочисленные подходы, в числе которых опрос родителей и родственников обследуемого, беседа с медицинским персоналом, воспитателями, анализ доступной медицинской документации. Важно, чтобы методика сбора информации была одинаковой для «случаев» и для контрольных наблюдений. В некоторых исследованиях можно добиться того, чтобы интервьюер или лицо, анализирующее медицинские документы, работало вслепую, не зная, кем является конкретный обследуемый («случай» это или контроль). В идеале интервьюер не должен быть осведомлен о рабочей гипотезе, проверяемой в ходе исследования, чтобы избежать индуцирования. При несоблюдении этих условий возможна ситуация, когда исследователь вольно или невольно будет с большей настойчивостью искать признаки предшествовавшего патогенного воздействия в ходе опроса лиц, страдающих заболеванием, или в ходе анализа медицинской документации, относящейся к таким лицам, что может существенно исказить результаты. В этой же связи особую ценность имеет анализ медицинских документов, заполнявшихся до того, как было диагностировано заболевание или зафиксирован неблагоприятный исход.

2.2.1.2. Когортные исследования

Когортные исследования (*follow-up study*) — вторая большая категория обсервационных аналитических исследований. Группа или группы индивидуумов определяются на основе наличия или отсутствия патогенных воздействий, являющихся предполагаемыми факторами риска заболеваний. Статус обследуемых оценивается в зависимости от их подверженности патогенному воздействию, на этот момент все потенциальные субъекты исследования не должны иметь признаков анализируемого заболевания. **Основная, или экспонированная (exposed), группа** — лица, подвергшиеся изучаемому воздействию. **Контрольная или неэкспонированная (unexposed), группа** — лица, не подвергавшиеся воздействию (Кельмансон И. А., 2002; Покровский В. П., Брико Н. И., 2012; Rothman К. J. [et al.], 2008). В дальнейшем за ними наблюдают в течение определенного периода времени и изучают частоту возникновения анализируемого заболевания в экспонированной и неэкспонированной группах.

Когортные исследования целесообразны для оценки ассоциации между воздействием и заболеванием (Кельмансон И. А., 2002; Rothman К. J. [et al.], 2008). Было бы неверным полагать, что когортные исследования являются более совершенными, чем исследования типа «случай — контроль»: каждое имеет свои сильные и слабые стороны, и они дополняют друг друга.

Достоинства когортных исследований:

— Участники исследования не имеют признаков заболевания на момент оценки их подверженности патогенному воздействию, а значит, не вызывает больших сомнений последовательность событий (воздействие — эффект) во времени.

— Когортные исследования пригодны для оценки эффекта, сопряженного с редко встречающимися воздействиями. Когортные исследования позволяют отобрать оптимальное число обследуемых лиц, испытавших на себе влияние этого воздействия и избежавших его.

— Когортные исследования дают возможность оценить множественные последствия одного патогенного воздействия.

Недостатки когортных исследований:

— Часто когортные исследования предполагают наблюдение за большим числом субъектов в течение многих лет, что сопряжено с существенными временными и материальными затратами. Вот почему когортные исследования обычно проводятся лишь после того, как удастся сформулировать рабочую гипотезу на основе предварительных данных исследований типа «случай — контроль».

— Когортные исследования подвержены влиянию ошибки, связанной с потерей части субъектов наблюдения с течением времени

(*losses to follow-up*), что тем более вероятно, чем длительнее период наблюдения, которой может составлять месяцы, годы, а иногда и десятилетия.

Когортные исследования могут быть как **проспективными**, так и **ретроспективными**, что зависит от взаимоотношения времени начала исследования и возникновения заболевания. Существует определенная терминологическая неточность, связанная с использованием понятий «ретроспективное исследование» и «проспективное исследование». Некоторые авторы используют термин «ретроспективное исследование» в качестве синонима исследований типа «случай — контроль», а термин «проспективное исследование» — как эквивалент когортного, поскольку исследования типа «случай — контроль» ориентированы от настоящего заболевания к прошлому, к изучению данных анамнеза с целью выявить возможную причину, тогда как когортные исследования направлены в будущее и ориентированы прежде всего на воздействие и только потом на развитие заболевания. Однако более правильно использовать термины «ретроспективное» и «проспективное» для обозначения того этапа, на который приходится начало исследования (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008). В исследовании по типу «случай — контроль» проводится отбор наблюдений на основе информации о наличии или отсутствии заболевания у обследуемых, а затем ретроспективно осуществляется сбор сведений о предполагаемом воздействии в предшествующий период. Следовательно, исследования типа «случай — контроль» являются ретроспективными. В любом варианте когортного исследования наблюдения выбираются на основе данных об имевшемся или отсутствовавшем патологическом воздействии, а в дальнейшем фиксируется развитие изучаемого заболевания; однако разница между ретроспективным и проспективным когортными исследованиями определяется тем, имелось ли заболевание на момент начала исследования (рис. 23).

— *Проспективное когортное исследование.* В начале исследования группы, состоящие из лиц, испытавших на себе влияние предполагаемого воздействия и избежавших этого воздействия, уже сформированы, но заболевание еще не развилось, что требует дальнейшего динамического наблюдения (*follow-up*) в течение достаточного периода времени.

В ходе проспективного исследования появляется возможность анализа более свежей медицинской документации, непосредственной оценки имевшегося воздействия, а также опроса участников исследования. Если численность анализируемой выборки достаточна, а последующее динамическое наблюдение полное, то при прочих равных условиях проспективное когортное исследование является более предпочтительным с точки зрения надежности и информативности.

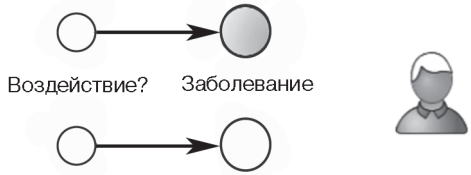
— *Ретроспективное когортное исследование.* Анализ начинается в то время, когда имеются сведения как об имевшемся патологическом воздействии, так и о развитии изучаемого заболевания. Исследователь формирует когорту задним числом и отслеживает судьбу ее членов до настоящего момента времени. Такой вариант также иногда именуется исследованием исторической когорты (*historical cohort*) (Altman D. G., 1999). Ретроспективное когортное исследование можно осуществить быстрее и с относительно меньшими затратами, так как все события уже произошли к моменту начала исследования. Оно особо оправдано при анализе заболеваний с длительным латентным периодом, при котором требуются годы наблюдений для реализации патогенного эффекта. Вместе с тем ретроспективные когортные исследования, базирующиеся на данных о наличии патогенного воздействия, имевшего место много лет назад, зависят от доступности достоверной информации из архивной медицинской документации. Поскольку документы, заполнявшиеся много лет назад, могли не предусматривать рассмотрение тех гипотез, которые являются предметом ретроспективного когортного исследования, могут возникнуть определенные трудности, связанные с неполнотой информации, содержащейся в этих документах, или ее несопоставимостью.

— *Двунаправленное когортное исследование.* Данные в одной и той же когорте получают как ретроспективно, так и проспективно (рис. 24). Подобный подход оправдан в тех ситуациях, когда патогенное воздействие может иметь как ближайшие, так и отдаленные по времени последствия.

Так, химические агенты могут иметь ближайшие последствия воздействия на популяцию, реализующиеся в течение нескольких лет, например, в виде увеличения риска рождения детей с задержкой внутриутробного развития, и отдаленные последствия, проявляющиеся спустя много лет, например, в виде повышенного риска синдрома дефицита внимания и гиперактивности у ребенка школьного возраста.

— *Гнездное когортное исследование (nested case-control design)* — дополнительная модификация когортного исследования, осуществляемого как проспективно, так и ретроспективно, с включением фрагмента исследования по типу «случай — контроль» (Кельмансон И. А., 2002; Hennekens C. H. [et al.], 1987; Rothman K. J. [et al.], 2008).

Исследование «случай – контроль»



Проспективное когортное исследование



Ретроспективное когортное исследование



Рис. 23. Различия между ретроспективными и проспективными исследованиями

Двунаправленное когортное исследование

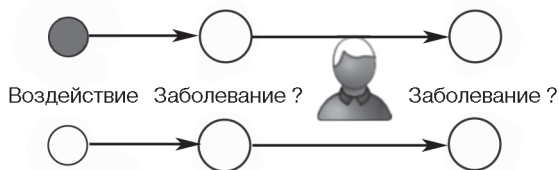


Рис. 24. Схема двунаправленного когортного исследования

Пример

Для оценки связи между уровнем липидов сыворотки крови у детей школьного возраста и риском развития тревожно-депрессивных расстройств во взрослом состоянии возможно исследование сыворотки крови с определением уровня липидов у когорты детей школьного возраста определенной численности. В дальнейшем необходимо исследование вероятности развития тревожно-депрессивных расстройств 10–20 лет спустя у этих лиц в зависимости от исходного уровня липидов сыворотки. В рамках гнездового исследования по принципу «случай — контроль» возможна следующая модификация когортного исследования: образцы сыворотки крови можно заморозить и направить на хранение, за обследуемыми лицами установить динамическое наблюдение. Если среди наблюдаемых субъектов будет зафиксировано достаточное число случаев развития тревожно-депрессивных расстройств во взрослом состоянии, образцы их сывороток крови могут быть подвергнуты анализу вместе с образцами сывороток контрольных наблюдений, сформированных из числа обследованных, не имеющих признаков тревожно-депрессивных расстройств. Таким образом, в данном примере гнездовое исследование по типу «случай — контроль», проводимое в рамках когортного проспективного исследования, позволяет подвергнуть анализу не все образцы сывороток, которых может насчитываться десяток тысяч, а лишь ограниченную их часть, что позволяет экономить ресурсы с учетом дороговизны биохимических исследований.

Проведение как проспективного, так и ретроспективного когортных исследований выдвигает ряд тактических задач.

Проблема оптимального отбора популяции, подвергшейся анализируемому воздействию. В зависимости от целей конкретного исследования и ряда практических соображений (распространен-

ности патогенного воздействия, необходимости получения полной информации об уровне воздействия и его последствиях для каждого обследуемого) определяются принципы ее формирования. В случае анализа относительно часто встречающихся патогенных воздействий, например курения, употребления алкоголя, не составляет большого труда идентифицировать достаточное число лиц, подвергшихся воздействию, опираясь на различные популяционные источники. Задача становится более сложной, если изучаются относительно редкие воздействия: идентификация лиц, подвергшихся воздействию, нередко проводится на основе данных, касающихся специальных популяций (например, лиц, проживающих в определенных географических и климатических условиях).

Использование специальных популяций (например, школьников, детей из домов ребенка) в качестве источника формирования обследуемой группы может быть продиктовано и необходимостью дальнейшего динамического наблюдения, так как с ними, возможно, проще установить систематический контакт в будущем. В подобных ситуациях выбор специальной группы не связан с большей подверженностью членов данной группы влиянию патогенного воздействия, и, следовательно, реализация такого подхода возможна лишь при достаточной распространенности изучаемого воздействия или большой численности группы.

Группа сравнения. Не менее ответственной задачей является формирование адекватной группы сравнения, состоящей из лиц, не подвергшихся воздействию (контрольной, неэкспонированной группы). В случае когортного исследования важнейший принцип ее формирования таков: сопоставляемые группы должны быть максимально схожи в отношении всех факторов, имеющих потенциальное отношение к риску развития заболевания, за исключением факта наличия или отсутствия конкретного изучаемого воздействия. При соблюдении этого принципа можно ожидать, что при отсутствии ассоциации между анализируемым воздействием и заболеванием частота развития заболевания в обеих сопоставляемых группах будет практически равной.

Важно также, чтобы информация, которую можно получить в группе лиц, избежавших патогенного воздействия, была сопоставима с информацией, доступной в группе лиц, подвергшихся этому воздействию. Это условие соблюдается в тех исследованиях, где объектом изучения является единая когорта, члены которой в дальнейшем делятся на отдельные группы в зависимости от их отношения к воздействию потенциального патогенного фактора. При этом создается внутренняя группа сравнения. Если в исследовании рассматривается влияние на риск развития заболевания сразу нескольких воздей-

твий, то в группу сравнения следует включать лиц, избежавших всех этих воздействий. Иначе говоря, предпочтительнее, чтобы группа сравнения представляла собой единую референтную популяцию, а не менялась по составу в зависимости от того, влияние какого фактора рассматривается в данный момент.

В случае когортных исследований, основанных на анализе группы лиц внутри специфической популяции, нередко вызывает затруднения формирование адекватной группы сопоставления из представителей той же популяции, так как не всегда можно с уверенностью сказать, в какой степени отдельные члены популяции испытали на себе влияние рассматриваемого патогенного воздействия. В частности, это может относиться к экологическим факторам. В таких случаях оправдано формирование внешней группы сравнения.

Пример

При оценке связи риска поведенческих нарушений у детей с загрязненностью воздуха продуктами химического производства возможно сопоставление частоты развития поведенческих нарушений у детей, проживающих в зоне экологического неблагополучия, с частотой аналогичных нарушений у детей того же возраста и пола, наблюдаемой в целом по региону или по стране.

Следует отметить, что подобный подход возможен лишь тогда, когда доступны общепопуляционные данные о частоте анализируемого заболевания (патологического состояния). Кроме того, использование общепопуляционных данных в качестве референтных предполагает, что лишь незначительная часть лиц из общей популяции могла быть подвержена тому патогенному воздействию, влияние которого является предметом анализа. Однако фактически определенная доля лиц, входящих в общую популяцию, может оказаться подверженной анализируемому патогенному воздействию (в примере с изучением влияния химического загрязнения атмосферы на риск нарушений поведения у детей, с аналогичными патогенными влияниями могут столкнуться некоторые дети из внешней группы сравнения, проживающие в других регионах). Вот почему использование общепопуляционной внешней группы сопоставления может привести к недооценке истинной ассоциации между патогенным воздействием и риском развития заболевания. Более того, использование общепопуляционных данных затруднено в связи с тем, что лица, формирующие такую группу сравнения, по ряду характеристик могут быть несопоставимы с представителями изучаемой когорты.

Несколько групп сравнения. Иногда в ходе когортных исследований представляется оправданным использование нескольких групп сравнения, особенно в тех случаях, когда ни одна из них в отдельнос-

ти не обеспечивает полного сходства с изучаемой группой, подвергшейся вероятному патогенному воздействию. При этом однотипность результатов сопоставлений с различными группами повышает уверенность исследователя в достоверности выводов.

Проблема медицинской документации как источника сведений. Существенным условием эффективного когортного исследования является доступность и полнота информации, позволяющей судить о том, в какой степени каждый субъект подвергался изучаемому воздействию и у кого из них развился интересующий исследователя исход. Сведения, относящиеся к наличию патогенного воздействия, дают многочисленные источники: медицинская документация, сведения, предоставленные непосредственно участниками в виде заполненных анкет и опросников; данные, полученные в ходе медицинского обследования или иных тестов; непосредственные оценки условий окружения, в которых пребывают обследуемые лица. Данные об исходах также могут быть получены на основе анализа медицинской документации, в частности историй болезни, амбулаторных карт, свидетельств о смерти, опросников, а также в результате непосредственного обследования лиц.

Ранее созданная медицинская документация в качестве источника информации содержит сведения о наличии или отсутствии предполагаемого патогенного воздействия и его уровне.

Достоинства медицинской документации предшествующего периода:

— Информация оказывается доступной для большинства членов обследуемой когорты.

— Документы заполняются до того, как планируется и начинается исследование, и на представленные в них сведения не может повлиять понимание целей и задач работы (Кельмансон И. А., 2002).

Недостатки медицинской документации предшествующего периода:

— В определенных ситуациях сведения, которые содержатся в архивных медицинских документах, не дают достаточной информации об уровне изучаемого патогенного воздействия, что может не удовлетворять задачам исследования.

— Возможно, отсутствуют данные о вероятных дополнительных факторах, учет которых также представляется целесообразным. Такого рода сведения часто могут быть получены лишь непосредственно у обследуемого лица или его родных, что заставляет прибегать к помощи опросников. Следовательно, возрастает вероятность некоторого искажения информации, связанной с субъективизмом оценок, особенно на фоне наличия у опрашиваемого лица изучаемого заболевания, и данное обстоятельство диктует особые требования к анкетированию с созданием максимально однотипных условий представления сведений каждым субъектом.

Для получения данных о ряде медико-биологических показателей (например, артериальном давлении, характеристиках электроэнцефалограммы) требуется непосредственное медицинское обследование, которое должно проводиться с соблюдением единых принципов и методик во избежание искажения данных. Если подобная информация необходима для оценки состояния здоровья пациента, то исследователь, осуществляющий медицинский осмотр или анализ лабораторных данных, в идеальном случае не должен быть осведомлен о том, испытывал ли данный конкретный субъект воздействие предполагаемого патогенного фактора.

Во многих когортных исследованиях предполагается исходное наличие единой классификации уровня предполагаемого патогенного воздействия в отношении каждого субъекта. Однако могут возникать ситуации, в которых с течением времени степень воздействия патогенного фактора на отдельных испытуемых меняется. Например, с течением времени могут меняться особенности диеты, социальные характеристики, бытовые условия и т. п. Игнорирование указанных изменений способно привести к недооценке истинной ассоциации между патогенным воздействием и его эффектом. Вот почему многие когортные исследования строятся таким образом, что предполагают периодическую ревизию статуса обследуемых лиц в зависимости от наличия или отсутствия влияния на них изучаемого фактора. При такой модификации исследования возможен учет общей длительности изучаемого воздействия, факта возникновения или устранения воздействия по ходу исследования, а также обстоятельств, которые могли повлиять на изменение статуса обследуемых лиц.

В любом варианте когортных исследований, как проспективных, так и ретроспективных, получение сведений об исходах предполагает динамическое наблюдение за членами когорты. Невозможность получить такую информацию у каждого испытуемого, а также непропорциональное представительство в динамическом наблюдении лиц, подвергшихся или не подвергшихся изучаемому воздействию, может существенно сказаться на полученных результатах. Значит, принципиальной задачей является сбор информации, касающейся каждого субъекта исследования, что и предопределяет чрезвычайную трудоемкость когортных исследований.

Продолжительность наблюдения за обследуемыми лицами будет во многом определяться латентным периодом развития заболевания, связанного с изучаемым патогенным фактором. Очевидно, чем продолжительнее предполагаемый латентный период (следовательно, и длительность наблюдения), тем более проблематичным становится осуществление когортного исследования, прежде всего проспективного, так как возрастает вероятность утраты связи

с обследуемыми. Если доля обследуемых лиц, выбывших из динамического наблюдения, не превышает 5 %, вероятность систематической ошибки невелика, если эта доля превышает 20 %, возникают серьезные сомнения в отношении значимости (валидности) результатов (Straus S. E., 2011). Еще более серьезная проблема, нежели просто высокая доля потерь в ходе динамического наблюдения, может возникать тогда, когда вероятность этих потерь зависит от самого факта наличия патогенного воздействия и (или) развития заболевания, связанного с этим воздействием. Например, с большей долей вероятности можно ожидать изменений адреса семей с неблагоприятным социальным статусом и, как следствие, утраты контакта с ними. Точно установить причину, приведшую к утрате контакта с обследуемыми лицами, не всегда возможно. Вот почему необходимо стремиться к тому, чтобы в ходе когортного исследования такие потери были сведены к минимуму. Кроме того, весьма важно сопоставить важнейшие характеристики тех обследованных, которые наблюдались систематически, и тех, которые были «потеряны» для исследования. Результаты исследования могут считаться валидными, если отсутствуют систематические различия между «оставшимися» и «утраченными» обследуемыми лицами по частоте выявления и выраженности у них важнейших потенциальных факторов риска заболевания и рассматриваемого патогенного воздействия. Важная проблема, связанная с «утратой» субъектов в ходе динамического наблюдения, состоит в том, что часть лиц, отобранных для исследования, может отказаться от дальнейшего участия в нем, руководствуясь мотивами, которые каким-то образом связаны с предполагаемым патогенным воздействием и (или) заболеванием. В подобных ситуациях отказ от участия в исследовании может серьезным образом повлиять на полученные результаты и возможность их обобщения. В случае связи мотивов отказа только с патогенным воздействием или только с возникновением заболевания выявляемая ассоциация, как правило, оказывается заниженной, и это дает основания полагать, что истинная выраженность ассоциации будет не меньше фактически выявленной. В то же время одновременная связь причин отказа от участия в исследовании с патогенным воздействием и с заболеванием может непредсказуемым образом отразиться на полученных результатах. Вот почему весьма желательно сопоставление важнейших характеристик тех лиц, которые согласились принять участие в исследовании, и тех, которые отказались от участия в нем. При этом особое внимание следует уделить сопоставлению частоты выявления и выраженности в указанных двух группах вероятных факторов риска заболевания и предполагаемого патогенного воздействия, а также самого заболевания. Отсутствие

систематических различий по указанным показателям позволяет доверять полученным результатам исследования.

Метод худшего и лучшего сценария (*worst case and best case scenario*). Косвенным методом, который может быть использован для оценки значимости потерь, является исследование силы ассоциации между патогенным воздействием и заболеванием при допущении экстремальных ситуаций. Суть метода сводится к допущению, согласно которому все субъекты, выпавшие из динамического наблюдения, либо были поражены изучаемым заболеванием либо, напротив, избежали этого заболевания (Straus S. E., 2011). Результаты таких расчетов дают тот диапазон значений, в котором будут находиться истинные показатели ассоциации.

Пример

Рассмотрим гипотетическую ситуацию: анализ частоты формирования синдрома дефицита внимания и гиперактивности в школьном возрасте среди 1000 детей, родившихся с признаками тяжелой внутриутробной гипоксии. Предположим, что 60 детей «выпало» из последующего динамического наблюдения, а диагноз «синдром дефицита внимания и гиперактивность» был зарегистрирован у 10 детей. Тогда фактический показатель заболеваемости составит:

$$10/940 = 10,6 \text{ на } 1000.$$

Согласно худшему сценарию, все 60 наблюдений, «потерянные» для исследования, должны быть отнесены к категории лиц с синдромом дефицита внимания и гиперактивности. В этом случае заболеваемость составит:

$$(10 + 60) / (940 + 60) = 70 \text{ на } 1000.$$

В соответствии с лучшим сценарием «потерянные» 60 наблюдений оказались здоровыми, в результате чего заболеваемость составит:

$$10/(940 + 60) = 10 \text{ на } 1000.$$

Таким образом, истинные значения колеблются в диапазоне от 10 до 70 на 1000; показатели, рассчитанные по лучшему сценарию, близки к вычисленному значению: 10,6 на 1000, однако показатель худшего сценария в 6,6 раз превышает расчетный. Результаты свидетельствуют о том, что «потеря» 60 наблюдений могла существенно повлиять на результаты исследования. Данный пример является иллюстрацией того, что большие «потери» неизбежно затрудняют интерпретацию результатов.

Цензурированные (*censored*) **наблюдения**. В тех ситуациях, когда за субъектом, входящим в когорту, не удается пронаблюдать до наступления некоторого события (например, до развития интересующего исследователя заболевания, патологического состояния, смерти или выздоровления, выписки из больницы, продолжительности грудного вскармливания и т. п.), говорят о цензурированных наблюдениях (Woodward M., 2014). В этом случае можно лишь сказать, что время до наступления события, интересующего исследователя, не меньше вре-

мени до последнего наблюдения. Время наступления события, таким образом, оказывается цензурированным (рис. 25).

Применение специальных статистических методов, учитывающих наличие цензурированных наблюдений, позволяет внести коррективы, связанные с неполнотой информации. Чаще всего с указанной целью используются методы **анализа выживаемости** (*survival analysis*). Методы получили свое название благодаря их изначально широкому применению в медицинских исследованиях для оценки продолжительности жизни при изучении эффективности методов лечения. Позднее данные методы стали применяться в социальных и психологических науках. В целом анализ выживаемости представляет собой построение моделей, описывающих данные о времени наступления того или иного события. Примером являются метод Каплана—Мейера, регрессия Кокса, принципы которых подробно рассматриваются в специальных руководствах (Altman D. G., 1999; Peat J. K., Barton B., 2014; Woodward M., 2014). Большинство прикладных статистических программ (SPSS, STATA, STATISTICA) позволяют осуществить подобные расчеты.

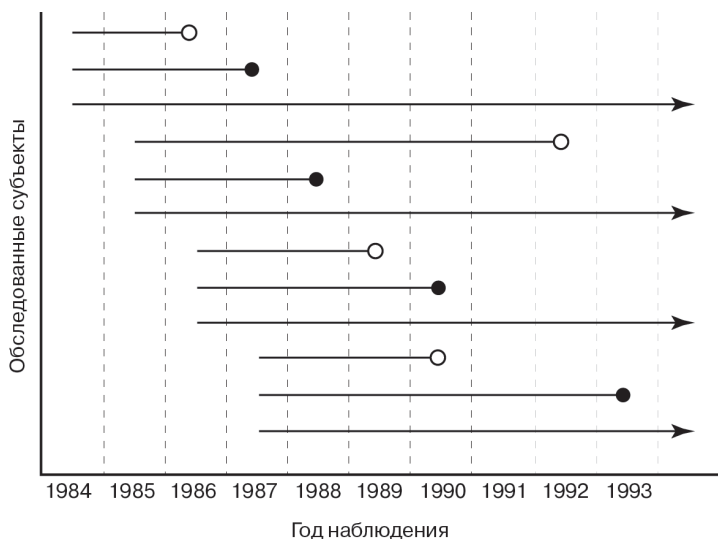


Рис. 25. Исследование с цензурированием данных:

темные кружки — субъекты, у которых зафиксирован изучаемый феномен (например, развития заболевания); светлые кружки — субъекты, выпавшие из дальнейшего наблюдения, у которых на момент последнего обследования заболевание отсутствовало (цензурированные наблюдения); стрелки — субъекты, у которых на момент последнего наблюдения отсутствовало заболевание, но за которыми возможно дальнейшее продолжение наблюдения (Woodward M., 2014)

2.2.2. Интервенционные исследования

Правильно спланированное **интервенционное исследование** в психологии является психологическим экспериментом, т. е. проводимым в специальных условиях опытом для получения новых научных знаний посредством целенаправленного вмешательства исследователя в жизнедеятельность испытуемого. Интервенционные исследования позволяют получить данные такого уровня значимости, которые сопоставимы с результатами исследований, применяемых в фундаментальных науках. В интервенционные исследования включаются лица в зависимости от их отношения к изучаемому воздействию, но только сам исследователь решает вопрос о том, кто из испытуемых подвергается этому воздействию, а кто — нет. Существенное преимущество, связанное с данным обстоятельством, состоит в возможности добиться достаточной численности обследуемой группы и, как следствие, валидности результатов. Экспериментатор вмешивается в естественный ход событий с тем, чтобы сконструировать ситуацию, позволяющую проверить теорию. Часто такая теория затрагивает вопросы причинно-следственных связей между отдельными переменными. Экспериментальный дизайн исследования представляет особый интерес для клинических психологов, поскольку любое терапевтическое вмешательство может рассматриваться как эксперимент.

В ходе экспериментального исследования используется ряд понятий.

Независимая, или экспериментальная, переменная — воздействие (лечение), варьируемое исследователем.

Зависимая переменная, или исход, — показатель, отражающий эффект экспериментального воздействия. **Конечный пункт (end-point)** — наступление некоторого события, вплоть до которого требуется наблюдение для регистрации исхода (Woodward M., 2014).

Статистические методы, чаще всего используемые в ходе проведения экспериментальных исследований, включают в себя дисперсионный анализ (анализ вариации, ANOVA) или связанный с ним *t*-критерий Стьюдента, многофакторный дисперсионный анализ (MANOVA), а также ковариационный анализ (ANCOVA) (Howell D. C., 2013; Peat J. K., Barton V., 2014; Woodward M., 2014).

В терапии и клинической фармакологии интервенционные исследования обычно обозначаются как клинические исследования (испытания) (Кельмансон И. А., 2002; Мелихов О. Г., 2003; Власов В. В., 2004; Покровский В. П., Брико Н. И., 2012). Применительно к лекарственным средствам (ЛС) и вакцинам клинические испытания классифицируют на 4 фазы:

— *Испытания I фазы* проводятся с привлечением малого числа добровольцев. Это могут быть здоровые добровольцы, иногда —

лица, страдающие заболеванием, для лечения которого разработано ЛС. Таких испытуемых обязательно приглашают в случае, если препарат обладает высокой токсичностью. Цель — выявление токсичности ЛС и его ожидаемого фармакологического действия.

— *Испытания II фазы* проводятся с участием небольшого числа испытуемых, имеющих состояние, которое должно быть объектом применения ЛС. Цель — выявление ожидаемого терапевтического эффекта, уточнение данных о безопасности и уточнение дозировки ЛС.

— *Испытания III фазы* проводятся на большом числе испытуемых, в последние десятилетия — с участием нескольких медицинских центров. Цель — выяснение размера терапевтического эффекта, уточнение данных о безопасности. Обычно на основании результатов испытаний III фазы ЛС разрешаются к применению.

— *Испытания IV фазы* проводятся в виде наблюдения за текущей практикой. Цель — уточнение размера эффекта от вмешательства в условиях реальной практики, уточнение данных о безопасности.

Интервенционные исследования могут быть терапевтическими (лечебными) и профилактическими (превентивными).

Терапевтические интервенционные исследования (исследования по вторичной профилактике) проводятся среди лиц, страдающих определенными заболеваниями, в целях изучения эффективности конкретного терапевтического подхода для уменьшения клинических проявлений, частоты обострений заболевания, снижения смертности.

Профилактические интервенционные исследования — вмешательства, применяемые по отношению к людям, у которых нет целевого состояния, с задачей предотвратить его развитие или при наличии состояния снизить вероятность неблагоприятных явлений.

Первичная профилактика — вмешательства при отсутствии патологического состояния. Исследования по первичной профилактике направлены на изучение влияния того или иного мероприятия с целью предупреждения возникновения заболевания или патологического состояния у лиц, не имевших признаков этого заболевания на момент начала исследования.

Вторичная профилактика — вмешательства, нацеленные на замедление развития заболевания и снижение вероятности осложнений.

Третичная профилактика — вмешательства, нацеленные на облегчение приспособления больного к вызванному болезнью дефекту, предотвращение инвалидности. Следовательно, профилактические исследования могут проводиться среди клинически здоровых лиц, характеризующихся как обычным, так и повышенным исходным риском развития заболевания.

Этический аспект. Активная роль исследователя в выборе тактики по отношению к каждому субъекту сопряжена с рядом этических соображений и необходимостью учета некоторых практических моментов, связанных с планированием и осуществлением работы. К числу важнейших этических аспектов относятся:

— недопустимость подвергать обследуемых лиц воздействиям с заведомо доказанными негативными эффектами. В подобных ситуациях допустимы интервенционные исследования, проводимые по принципу «от обратного»: изучается эффект элиминации (устранения) негативного воздействия. Например, возможно проведение интервенционного исследования, направленного на изучение эффекта прекращения курения матери во время беременности с точки зрения риска рождения ребенка с малой массой тела (Кельмансон И. А., 2002);

— недопустимость лишения в рамках исследования части пациентов тех лечебных воздействий, в которых они могут нуждаться в связи с заболеванием.

Практическими аспектами, определяющими осуществимость интервенционных исследований, являются возможность обеспечения достаточной численности обследуемых и решение задач материально-технического обеспечения. Можно без преувеличения сказать, что четко спланированные и осуществленные интервенционные исследования являются своеобразным эталоном исследований, а результаты таких исследований и выводы, как правило, вызывают максимальное доверие.

Планирование интервенционного исследования предполагает решение ряда вопросов:

- выбор обследуемой популяции;
- распределение обследуемых в зависимости от назначения им изучаемого интервенционного воздействия;
- обеспечение соблюдения участниками предписанного режима;
- стандартизация методики оценки результатов.

Группа обследуемых, входящих в интервенционное исследование, формируется в иерархической последовательности (рис. 26).

Референтная популяция — группа лиц, в отношении которой представляется допустимым проведение соответствующего интервенционного исследования. Референтная популяция может не иметь каких-то специальных ограничений или быть ограничена **критериями включения и исключения** (*inclusion criteria, exclusion criteria*): лица определенного возраста, пола, географические или социальные рамки, которые, по мнению исследователя, могут существенным образом модифицировать результаты. Ясные критерии включения пациентов (обследуемых) необходимы для описания популяции (генеральной сово-

купности), которой соответствуют включенные в исследование субъекты. Так, если исследуются пациенты с депрессивными расстройствами, то должно быть определено, по какому критерию устанавливается наличие этих расстройств (например, критерии Американской психологической ассоциации). Если обследуются поликлиники, то необходимо уточнить расположение (в селе или в городе), размер, штат врачей, численность обслуживаемого населения. Ясные критерии исключения необходимо сформулировать, чтобы сделать изучаемую группу пациентов достаточно однородной. Например, из числа участников эксперимента исключают лиц с тяжелыми сопутствующими болезнями, с состояниями, мешающими выполнить условия эксперимента (например, с деменцией), тех, кому вмешательство может быть вредно. Исключение нетипичных случаев обеспечивает однородность группы пациентов, т. е. меньшую вариабельность переменных в исходном состоянии и в оценке величины эффекта вмешательства (Власов В. В.,

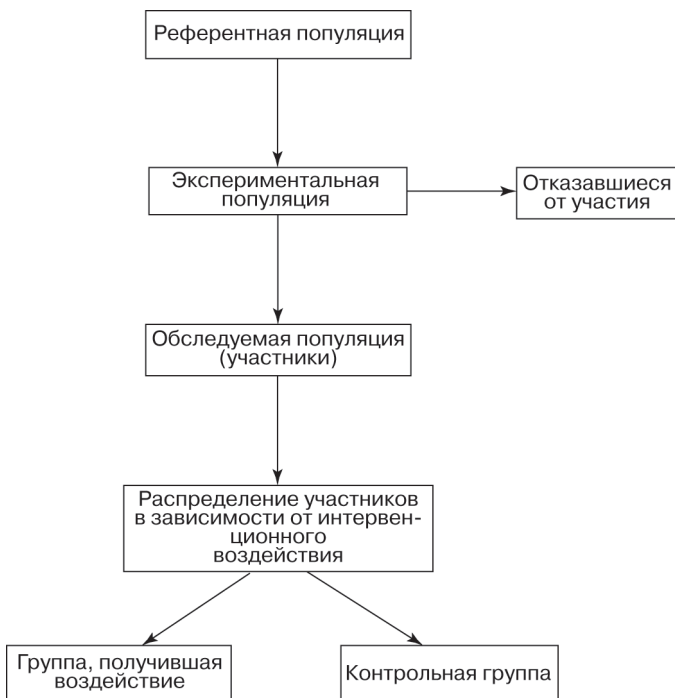


Рис. 26. Этапы формирования групп, входящих в интервенционное исследование

2004). Таким образом, увеличивается статистическая чувствительность эксперимента. В целом выбор референтной популяции предопределяет последующую обобщаемость результатов, полученных в ходе исследования.

В исследование могут быть включены только те испытуемые, для которых изучаемые вмешательства могут принести сравнимую пользу. Кроме того, исследования проводятся в тех ситуациях, когда отличие эффекта испытываемого вмешательства от эффекта плацебо не было ранее установлено (Власов В. В., 2004).

Экспериментальная популяция представляет собой реальную группу индивидуумов, среди которых проводится испытание. Оптимально, чтобы экспериментальная популяция существенно не отличалась от референтной, что позволит обобщить полученные результаты в отношении референтной популяции. Важно учесть, достаточна ли предполагаемая численность экспериментальной популяции для обеспечения достоверности результатов. Часто с целью увеличить численность экспериментальной популяции прибегают к многоцентровым (*multicenter*) исследованиям, объединяющим данные многочисленных медицинских учреждений в пределах отдельного региона, страны, а нередко и различных стран. Не менее важно отобрать экспериментальную популяцию так, чтобы обеспечить сбор окончательной информации в объеме, адекватно отражающем результаты различных вариантов воздействий на достаточное число испытуемых и в течение достаточно продолжительного периода наблюдения. После отбора экспериментальной популяции предполагаемых участников приглашают к участию в исследовании. В обязательном порядке им предоставляют полную информацию о целях исследования, предполагаемых процедурах и связанных с ними возможных негативных эффектах и положительных результатах. Очевидно, в случае экспериментальной детской популяции такая информация должна быть предоставлена родителям или опекунам.

В случае получения согласия на участие в исследовании (информированного согласия, *informed consent*) предполагаемые участники должны быть обследованы на предмет того, в полной ли мере они удовлетворяют заранее заданным критериям включения и исключения. Основаниями для исключения того или иного субъекта из исследования могут быть его потребности в специфическом лечении или, напротив, индивидуальные противопоказания в отношении отдельных терапевтических воздействий, планирующихся в ходе исследования, а также недавнее участие в ином интервенционном исследовании.

Обследуемая популяция (*study population* — лица, прошедшие отбор (желающие участвовать в исследовании и удовлетворяющие

заданным критериям), непосредственные участники исследования. Нередко обследуемая популяция составляет лишь небольшую группу в рамках экспериментальной популяции. Важно помнить о том, что лица, вошедшие в обследуемую популяцию, по ряду характеристик могут существенно отличаться от членов экспериментальной популяции. Фактор волонтерства или, напротив, отказа от участия в исследовании во многом может быть продиктован социально-экономическим статусом семьи, возрастом и образованием, а также влиянием иных, нередко трудно учитываемых факторов, многие из которых могут самостоятельно влиять на риск заболеваемости и смертности.

Отклик (*response rate*), т. е. доля людей, откликнувшихся на просьбу включиться в исследование, соответствующих критериям включения и исключения, должен быть достаточно высок: не менее 80 %. В противном случае сформированная обследуемая группа, возможно, не будет соответствовать сформулированным критериям включения и исключения. Возможно, результат испытания, от участия в котором отказалось много пациентов, будет неприложим ко всем людям (например, к тем, кто отказывается). Так, пациенты, активно выбирающие некоторый метод лечения, лечатся старательнее, правильнее выполняют назначения, чаще общаются с исследователем и сохраняют с ним контакт. Это свойство часто именуют **комплайенсом** (*compliance*), под которым понимают исполнительность, аккуратность, склонность к сотрудничеству, податливость, согласие с режимом, подчинение указаниям.

Численность. Не менее пристального внимания заслуживает вопрос об адекватной численности обследуемой группы. Эта проблема актуальна в исследованиях любого типа, однако если в наблюдательных исследованиях предметом изучения являются выраженные эффекты, что позволяет ориентироваться на группы меньшей численности, в интервенционных исследованиях для достижения адекватной статистической силы, позволяющей зафиксировать даже умеренные эффекты, требуется анализ относительно многочисленных групп. Даже при уверенности в том, что интервенционное воздействие может оказать существенное влияние на изучаемый показатель (например, его снижение на 50 % и более), разумнее планировать исследование в расчете на более умеренный эффект (например, снижение или увеличение показателя на 10–20 %) и закончить его досрочно в случае выявления более оптимистичных промежуточных результатов.

Важнейшими факторами, влияющими на силу интервенционного исследования, следует считать **число «конечных пунктов»** (*endpoints*), достигнутых участниками исследования (т. е. число участников, прошедших исследование от начала до конца), а также нали-

чие различий в выполнении условий исследования лицами, вошедшими в сопоставляемые группы. Для достижения достаточного числа «конечных пунктов» необходимо обеспечить:

- участие в исследовании лиц из популяции высокого риска;
- адекватную продолжительность динамического наблюдения;

Необходимо контролировать одинаковое соблюдение всеми участниками исследования условий протокола.

Испытательный срок. Если лица, вошедшие в группу интервенционного воздействия, не выполняют соответствующих предписаний, или, напротив, лица из контрольной группы наблюдений в определенной степени подвергаются влиянию того же интервенционного воздействия, что и пациенты из группы клинического испытания (например, самостоятельно применяют аналогичные лекарственные препараты, сходным образом меняют режим, диету и т. п.), конечные результаты исследования могут быть сильно искажены. В целях повышения дисциплинированности участников нередко перед непосредственным началом исследования искусственно создают период отсева, своеобразный испытательный срок продолжительностью несколько недель или месяцев, в течение которого потенциальные участники еще до осуществления формальной рандомизации получают интервенционное воздействие или плацебо, и те, кто испытывают трудности с соблюдением протокола, в дальнейшем исключаются из исследования. Такая тактика особо оправдана, когда отсутствует необходимость проводить интервенционное вмешательство сразу после какого-то критического события (например, начала заболевания). Зачастую предпочтительно уменьшить общую численность наблюдений, ограничившись лишь дисциплинированными участниками, чем проводить испытание в группе большей численности, состоящей из лиц, плохо соблюдающих условия. Практическая реализация идеи испытательного срока для вероятных участников исследования зависит от конкретной ситуации. В случае испытания нового лекарственного препарата оптимальным можно считать назначение на этой стадии плацебо всем предполагаемым участникам исследования; предпочтительно, чтобы факт приема плацебо мог быть проконтролирован, например, за счет учета его метаболитов в моче. Неявка пациентов на контрольные врачебные осмотры, а также признаки того, что пациент не принимает плацебо в соответствии с рекомендациями, могут быть основанием для его исключения из дальнейших исследований. Возражения против идеи испытательного срока сводятся к тому, что факторы, повлиявшие на большую или меньшую дисциплинированность потенциальных испытуемых, могут быть сами по себе связаны с вероятностью реализации изучаемого эффекта. С одной стороны, подтверждение или опровержение такой возможности реально в

том случае, если в ходе исследования прослеживается катамнез тех пациентов, которые были исключены из клинического испытания. С другой стороны, факт возможной связи дисциплинированности испытуемых с риском развития изучаемого исхода может иметь значение лишь для вывода о возможности обобщить полученные результаты к популяции в целом (*внешней валидности*), но не влияет на значимость результатов внутри изученной группы (*внутреннюю валидность*). Поскольку задача обеспечения внутренней валидности является первоочередной, идею испытательного срока можно считать оправданной.

По определению, интервенционное исследование требует согласия и активного сотрудничества испытуемых. Однако после высказанного согласия на участие в исследовании часть испытуемых может в силу тех или иных причин все-таки уклониться от предписанного протокола. Поводом может послужить появление побочных эффектов, невозможность регулярного приема лекарства или простое нежелание. Нередко само по себе состояние пациента диктует необходимость изменения терапевтической тактики. Вот почему важнейшими задачами являются контакт с участниками исследования, реализация мер, направленных на обеспечение максимального сотрудничества испытуемых, соблюдение ими условий протокола. Вероятность невыполнения условий сильно зависит от продолжительности исследования, сложности и строгости протокола. Возможны различные подходы, направленные на обеспечение максимального соблюдения участниками требований протокола. Прежде всего, имеет значение критический отбор участников, так как их заинтересованность способствует большему соблюдению поставленных условий. Если исследование направлено на изучение эффективности мер профилактики конкретного заболевания, отбор для участия лиц, характеризующихся повышенным риском развития этого заболевания, является сильным фактором мотивации. Для обеспечения участия в испытании необходим постоянный контакт, с этой целью исследователь проводит беседы с участниками по телефону, посещает их на дому, регулярно проводит амбулаторные осмотры. Повысить мотивацию обследуемых можно и предоставлением им той медицинской информации, которую они не смогли бы почерпнуть из других источников. Большое значение имеет и соблюдение поставленных условий исследования, что зачастую трудно объективно оценить. В некоторых программах интервенционных исследований, предполагающих изменения режима и поведения, единственным источником для суждения о соблюдении условий протокола является отчет самого обследуемого. Фармакологические интервенционные исследования располагают большим арсеналом средств контроля выполнения условий: подсчет числа использован-

ных таблеток, определение уровня метаболитов или специальных меток в моче. Следует отдавать себе отчет в том, что невыполнение условий протокола некоторыми испытуемыми — неизбежная ситуация. Однако чрезвычайно важно продолжать динамическое наблюдение даже за теми лицами, которые изначально вошли в исследование, а затем отказались от соблюдения его условий. Независимо от того, в какой мере обследуемая популяция соответствует экспериментальной популяции в целом, возможные различия не могут повлиять на результаты сопоставлений внутри самой группы участников исследования; имеющиеся несоответствия могут лишь отразиться на возможности обобщения полученных результатов в отношении всей экспериментальной или референтной популяции. С учетом сказанного представляется принципиально важным сопоставление важнейших характеристик лиц, согласившихся на участие в исследовании и отказавшихся от участия. В некоторых исследованиях, где ожидается множество отказов от участия в них, применяется следующий прием: пациентам, которые на участие в исследовании не согласились, предоставляется то традиционное лечение, которое им показано по заболеванию, а затем результаты лечения у них сравнивают с результатами в группах, составивших обследуемую популяцию.

Задача, которую необходимо решить после определения участников исследования, — это решение вопроса о том, кто и как получит изучаемое интервенционное воздействие.

2.3. Нерандомизированный (квазиэкспериментальный) и рандомизированный дизайны

Нерандомизированный и рандомизированный дизайны исследования можно выделить в зависимости от подходов к организации интервенционного исследования (Barker C. [et al.], 2002). Нерандомизированный дизайн также именуется некоторыми авторами как квазиэкспериментальный (Cook T. D., Campbell D. T., 1979).

2.3.1. Нерандомизированный (квазиэкспериментальный) дизайн

Дизайн по принципу «одна группа, только последующее тестирование» может быть схематично представлен следующим образом:

$$XO,$$

где X — экспериментальное воздействие (лечение); O — исследованный показатель или показатели.

Отсутствует изучение исходных показателей у лиц, подвергшихся воздействию, показатель оценивается только после воздействия

и лишь в одной группе испытуемых. По сути, такое исследование ближе к описательным, чем к аналитическим. Указанный дизайн исследования приемлем, например, при изучении удовлетворенности пациентов качеством оказанной им медико-психологической помощи. В этом случае O будет представлять собой уровень удовлетворенности пациента. Данный дизайн может быть полезным для формулировки гипотезы, однако не позволяет говорить о каких-либо причинно-следственных связях. Попытка усмотреть причинно-следственную связь в данном случае соответствовала бы классической ошибке при оценке каузальности: *post hoc ergo propter hoc* («после того, следовательно, вследствие того») (Barker C. [et al.], 2002; Damer T. E., 2009).

Дизайн по принципу «одна группа, предварительное тестирование и последующее тестирование» можно представить следующей схемой:

$$O_1 X O_2.$$

Дизайн является продолжением предыдущего с той лишь разницей, что добавляется предварительное тестирование, и это позволяет сопоставить его с конечными данными, полученными спустя какое-то время после интервенционного воздействия в ходе последующего тестирования. Например, следуя указанному дизайну, можно изучить исходную выраженность симптомов депрессии у пациентов, сопоставить с ней показатели, выявленные у всех обследованных после проведения психотерапевтического лечения. Однако не всегда представляется возможным и обоснованным однозначно связать изменения показателей теста с интервенционным воздействием X . Как и в предыдущем случае, существует вероятность логической ошибки: *post hoc ergo propter hoc*. Дизайн имеет недостатки, негативно влияющие на его внутреннюю валидность, т. е. возможность дать ответ на вопрос, связаны ли фиксируемые изменения показателей с изучаемым воздействием:

— *Вероятность эндогенных изменений*, возникающих у обследованных лиц, независимо от внешних факторов, в том числе изучаемых интервенционных воздействий, в частности, возможна спонтанная ремиссия заболевания.

— *Возрастные изменения*, связанные с ростом и созреванием субъекта (относятся к эндогенным изменениям и особо актуальны в детском возрасте: нередко с возрастом дети избавляются от определенных транзиторных нарушений).

— *Реактивность измерения*: под влиянием процесса измерения меняется сам измеряемый показатель. Например, при психологическом тестировании испытуемые нередко лучше справляются с тестом при его

повторном предъявлении. Кроме того, детальное интервьюирование и тщательный сбор анамнеза при предварительном тестировании могут способствовать более глубокому пониманию клинической проблемы у обследуемых лиц и лучшему клиническому исходу.

— *«Эпохальный сдвиг»*, который может способствовать в целом изменению привычек, стиля поведения и т. п. Например, благодаря пропаганде отказ от курения становится признаком хорошего тона и обследуемые могут отказываться от курения с течением времени независимо от проводимого интервенционного вмешательства (например, психотренинга).

— *Влияние сторонних факторов*, т. е. значимых событий, отличных от интервенционных вмешательств, имевших место на интервале между предварительным тестированием и последующим тестированием. Например, общий экономический кризис, разразившийся за это время, способствует повышению уровня тревоги у испытуемых; повышение цен на табачную продукцию — отказу от курения.

— *Регрессия к среднему*. Участники клинических испытаний часто отбираются на основе выявления у них экстремальных показателей тестов (очень высоких или очень низких), свидетельствующих о наличии у них той или иной патологии. Феномен регрессии к среднему связан с ненадежностью теста, и это означает, что при повторном тестировании испытуемого показатели будут «сдвигаться» в область, ближе к средним (исходно высокие — в сторону понижения, и наоборот). Иными словами, будет создаваться впечатление улучшения показателей, даже при неэффективности терапии (Cook T. D., Campbell D. T., 1979).

Не менее существенной проблемой, связанной с указанным дизайном исследования, является **конструктивная валидность** интервенционного воздействия. Оценка конструктивной валидности предполагает, что изучаемое вмешательство вызывает реальные изменения (т. е. обладает внутренней валидностью), однако ставит вопрос: что именно в этом вмешательстве (какой конструкт) фактически способствует возникновению изменений? К числу важнейших проблем, связанных с конструктивной валидностью, можно отнести следующие:

— *искажающие (смещающие) переменные, конфаундеры (confounding variables)*. Термин «искажающие» означает, что эти переменные влияют одновременно с интервенционным воздействием и неразрывно с ним связаны. В качестве искажающих переменных могут выступать такие факторы, как личностные характеристики психолога, его профессиональные качества и т. п.;

— *эффект ожидания*, когда положительный терапевтический эффект может быть связан просто с тем, что пациенту уделяют внимание, а не с конкретным интервенционным воздействием, эффектив-

ность которого оценивается. Эффект ожидания проявляется в форме плацебо-эффекта в случае клинических испытаний лекарственных средств;

— *эффект Хоторна (Hawthorne effect)*: само по себе проведение исследования оказывает положительное влияние на пациента. Условия новизны и интереса к эксперименту, повышенное внимание к самому исследованию приводят к положительным результатам, что является искажением реального положения вещей (Cook T. D., Campbell D. T., 1979).

Выявляемые различия между показателями O_1 и O_2 (т. е. общая разница между предварительным и последующим тестированием) иногда именуется **брутто-эффектом интервенционного воздействия** (Rossi P. H. [et al.], 2004). При этом нетто-эффект определяется как эффект, который обоснованно может быть связан непосредственно с интервенционным воздействием, и представляет собой разность между брутто-эффектом и эффектом, связанным с влиянием искажающих переменных и ошибки измерения.

Нередко дизайн $O_1 X O_2$, обозначаемый также как открытое клиническое испытание, применяется в качестве предварительного этапа, направленный на выявление наличия брутто-эффекта как такового. В дальнейшем могут использоваться более сложные дизайны исследования для выявления нетто-эффекта.

Дизайн по принципу «неэквивалентные группы, только последующий тест» может быть представлен следующей схемой:

$$\begin{array}{l} NR X O, \\ NR O, \end{array}$$

где NR — отнесение испытуемых к сопоставляемым группам без рандомизации.

Подобный дизайн напоминает вариант «одна группа, только последующий тест», однако отличие состоит в том, что группа, подвергнутая экспериментальному воздействию X , сопоставляется с группой, которая такого воздействия не получает. Например, интервенционное воздействие организуется для пациентов в одной больничной палате и не проводится у пациентов в другой палате, последние будут выступать в качестве группы сопоставления (контрольной группы). Возможна модификация указанного дизайна: пациенты будут получать активное лечение различных видов, например три вида терапии, назначаемой пациентам в трех различных палатах. Данный дизайн исследования не позволяет непосредственно сопоставить изменения показателей последующего и предварительного тестирования. Вот почему такой дизайн возможен при ретроспективных исследованиях, когда отсутствует информация об исходных показателях,

и информация о состоянии пациентов может быть получена лишь в результате последующего тестирования. По своему смыслу такой дизайн приближается к дизайну корреляционного наблюдательного исследования, так как исследование сводится к выявлению ассоциации между принадлежностью пациента к одной из групп и величиной показателя, характеризующего эффект лечения. Серьезной угрозой внутренней валидности подобного дизайна является неконтролируемый отбор испытуемых: поскольку отнесение участников исследования к сопоставляемым группам не является рандомизированным, нет уверенности в том, что пациенты, оказавшиеся в этих группах, имели равные стартовые условия до начала воздействия X. Пациенты из сопоставляемых групп могут систематически различаться между собой по ряду признаков, например по уровню мотивации, степени тяжести заболевания, демографическим характеристикам. Даже если исследователь может сопоставить группы по указанным показателям и не выявить существенных различий, всегда существует вероятность наличия каких-то неучтенных важных систематических различий.

Дизайн по принципу «неэквивалентные группы, предварительное тестирование, последующее тестирование» может быть схематично представлен следующим образом:

$$\begin{array}{c} NR O X O, \\ NR O (Y) O. \end{array}$$

Дизайн используется весьма часто и содержит некоторые черты двух рассмотренных ранее дизайнов. Вторая группа в нижней части схемы, которая не получает экспериментального воздействия, именуется контрольной группой. Возможны различные подходы к ее формированию:

- назначение пациентам из контрольной группы альтернативно-го интервенционного воздействия (Y на приведенной схеме);
- отсутствие какого-либо воздействия вообще.

Подход к формированию контрольной группы зависит от постановки задач исследования: попытка выявить преимущество одной терапевтической схемы по сравнению с другой или попытка доказать, что терапевтическое воздействие эффективнее, чем его отсутствие. Основная угроза внутренней валидности связана с неконтролируемым отбором испытуемых: лица, вошедшие в сопоставляемые группы, могут систематически различаться между собой по ряду важных характеристик, а не только по наличию или отсутствию интервенционного воздействия. В ряде случаев исследователи пытаются учесть и нивелировать подобные различия, используя специальные статистические методы: множественный регрессионный анализ, ковариационный анализ (Cook T. D., Campbell D. T., 1979; Shadish W. R.

[et al.], 2001). Например, если пациенты из основной группы оказываются моложе, чем пациенты из контрольной группы, возраст пациента может использоваться как ковариационный признак в ходе статистического анализа, что позволяет изучить парциальный эффект ковариационного признака, в данном случае возраста. Однако такие статистические приемы могут оказаться неэффективными, если неэквивалентные группы сформированы из представителей различных популяций: ситуация напоминает сопоставление слона и мыши с поправкой на массу этих животных (Barker С. [et al.], 2002). Чтобы минимизировать риск неэквивалентности групп, можно попытаться подобрать испытуемых в обеих группах с учетом важнейших, ключевых показателей, таких как возраст, пол, степень тяжести заболевания и т. п., т. е. использовать принцип матчинга. Подобный подход иногда именуется **конструированными контрольными наблюдениями** (*constructed controls*) (Rossi P. H. [et al.], 2004). Однако практически реализовать такую задачу весьма затруднительно; кроме того, по-прежнему существует вероятность того, что пациенты будут представлять различные популяции.

2.3.2. Рандомизированный дизайн

Рандомизированный экспериментальный дизайн отличается тем, что распределение участников исследования по группам, в которых условия эксперимента различны, основано на принципе рандомизации. Рандомизированный отбор предполагает, что каждый из обследуемых субъектов имеет равные шансы быть участником той или иной интервенционной программы, и выбор тактики в отношении одного обследуемого не зависит от тактического решения в отношении другого. Реализация указанного принципа может осуществляться таким образом, что число обследуемых лиц, подвергаемых изучаемому воздействию, окажется случайным либо, в случае заранее оговоренной численности, случайно формируются сопоставляемые группы. Чаще всего для достижения рандомизации используют таблицу случайных чисел или генерирование списка случайных чисел с использованием прикладных статистических программ, что рассматривалось в предыдущих разделах (см. Исследование популяционных срезов). В простейшем случае для распределения обследуемых лиц в зависимости от подверженности изучаемому воздействию с использованием таблицы случайных чисел поступают следующим образом: если имеется N потенциальных испытуемых, из которых n должно быть подвержено воздействию, то всех испытуемых нумеруют числами от 1 до N ; лица, списочные номера которых соответствуют первым n числам, полученным по таблице случайных чисел, должны войти в группу,

подвергаемую изучаемому интервенционному воздействию, а остальные $N - n$ обследуемых составят контрольную группу наблюдений.

В ряде случаев могут быть основания предполагать, что исход воздействия зависит от ряда дополнительных характеристик, таких как пол обследуемого, стадия или клиническая форма заболевания, и тогда необходимо, чтобы сопоставляемые группы были сбалансированы и по этим показателям. Достигнуть такой цели можно при помощи более сложной процедуры рандомизации, именуемой **методом блоков** (*blocking*), когда каждый участник исследования классифицируется в зависимости от рассматриваемых показателей, что приводит к формированию отдельных подгрупп (блоков), а затем осуществляется рандомизация отдельно внутри каждого блока. Например, если имеется два блока и необходимо осуществить отбор n испытуемых для предполагаемого интервенционного воздействия, то внутри каждого блока осуществляют отбор $n/2$ испытуемых.

Важное достоинство принципа рандомизации состоит в том, что минимизируется вероятность ошибки отбора и повышается внутренняя валидность эксперимента. Соблюдение принципа рандомизации позволяет исследователю манипулировать на каждом этапе эксперимента какой-то одной переменной, вследствие чего ассоциация, выявляемая между воздействием и эффектом, с большей вероятностью отражает истинную причинно-следственную связь.

Рандомизированные контролируемые испытания (РКИ) (*randomized controlled trials, RCT*) — рандомизированные экспериментальные исследования, прежде всего для оценки эффективности лекарственных средств, широко распространены в медицине. В клинической психологии РКИ часто обозначаются как исследования эффекта (*efficacy studies*), т. е. доказательств того, что в идеальных условиях интервенционное воздействие работает, вызывает больше пользы, чем вреда. Этим они отличаются от прагматичных исследований эффективности (*effectiveness studies*), которые направлены на выявление того, как в целом в обычных условиях лечение влияет на пациента; чаще всего это неконтролируемые исследования, проводимые в полевых условиях (Barker С. [et al.], 2002).

Рандомизированные группы, предварительное тестирование и последующее тестирование — наиболее распространенный вариант рандомизированного экспериментального дизайна, который может быть описан следующей схемой:

$$\begin{array}{c} R O X O, \\ R O (Y) O, \end{array}$$

где R — рандомизированное отнесение обследуемых лиц к различным условиям эксперимента.

Процедура рандомизации может быть осуществлена при помощи таблицы случайных чисел, генератора случайных чисел. Многие прикладные статистические программы также позволяют распределить испытуемых в зависимости от их отношения к интервенционному воздействию на основе рандомизированного подхода. Например, такую процедуру можно выполнить при помощи программы WINPEPI (Abramson J. H., 2011). **Межгрупповой фактор** (*between-groups factor*) — независимая переменная, т. е. признак, говорящий о том, получают участники экспериментальное интервенционное вмешательство или нет. Именно на его основании участники исследования делятся на группы. В исследовании, соответствующем приведенной схеме, межгрупповой фактор имеет два уровня (основная и контрольная группа). Однако может быть больше одной основной и больше одной контрольной группы. Например, исследование может изучать эффекты психодинамической терапии, поведенческой терапии и сопоставлять их с психологическими характеристиками у лиц, находящихся в листе ожидания, т. е. не получающих лечение (контрольная группа). В этом случае межгрупповой фактор имеет 3 уровня (Sloane R. B., 1975).

Многофакторный дизайн (*multi-factorial design*) — дизайн, имеющий больше одного межгруппового фактора (рис. 27). Изучение влияния какого-то одного фактора является простейшим подходом к осуществлению интервенционных исследований. Однако с практической точки зрения нередко представляется оправданной проверка в ходе одного исследования сразу нескольких рабочих гипотез. Так, исследование, предполагающее анализ двух рабочих гипотез, может основываться на многофакторном подходе вида 2×2 , при котором участники исследования сначала распределяются на основе принципа рандомизации в зависимости от терапии по схеме α или β , что направлено на проверку первой гипотезы, а затем в пределах каждой группы осуществляется дополнительное распределение на основе принципа рандомизации в отношении проведения терапии по схеме А или В, что направлено на проверку второй гипотезы.

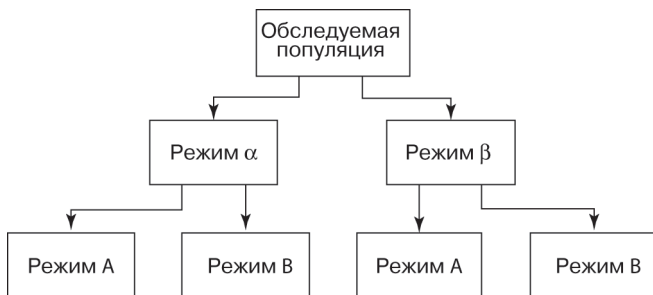


Рис. 27. Многофакторный дизайн экспериментального исследования

Пример

В исследовании «Шеффилдский психотерапевтический проект» (Sheffield Psychotherapy Project) имелось два межгрупповых фактора: двухуровневый фактор, определяющий психотерапевтический подход (когнитивно-бихевиоральная или психодинамически-интерперсональная терапия), и двухуровневый фактор, определяющий продолжительность терапии (8 или 16 сессий) (Shapiro D. A., 1996).

Возможен многофакторный подход по принципу $2 \times 2 \times 2$, при котором оценке подлежат 3 гипотезы, а обследуемые субъекты распределяются в зависимости от 3 режимов терапии; допустимо и многофакторное планирование исследований более высокого порядка.

Преимуществом многофакторного планирования интервенционных исследований является возможность изучения независимого влияния отдельных рассматриваемых факторов и их взаимодействия. С указанной целью часто прибегают к многофакторному дисперсионному анализу (MANOVA) (Altman D. G., 1999; Howell D. C., 2013; Peat J. K., Barton B., 2014).

Дизайн с предварительным тестированием и последующим тестированием является одним из вариантов **дизайна с повторными измерениями** (*repeated-measures design*), т. е. такого, при котором каждый участник обследуется в два момента времени и более. Число таких моментов, в которые проводится обследование участников эксперимента, может быть и большим, например каждые 6 мес., каждый год после окончания интервенционного воздействия и т. п. Статистический анализ в подобных ситуациях чаще всего выполняется при помощи дисперсионного анализа с повторными измерениями (*repeated measures ANOVA*) (Peat J. K., Barton B., 2014).

В тех случаях, когда исследование предполагает наличие многих промежуточных моментов, необходимо предусмотреть возможность возникновения такой ситуации, когда потребуется модификация протокола исследования или даже прекращение последнего раньше намеченных сроков. Для обеспечения благополучия пациентов необходимо контролировать промежуточные результаты, причем желательно, чтобы эту функцию выполняли независимые эксперты. С одной стороны, если промежуточные результаты свидетельствуют о безусловной пользе или очевидных негативных последствиях проводимого интервенционного воздействия, есть все основания для прекращения исследования. В подобных ситуациях нецелесообразно и неэтично как воздерживаться от очевидно эффективного воздействия в отношении лиц из контрольной группы наблюдений, так и подвергать испытуемых заведомо вредному воздействию. С другой стороны, нужно избегать соблазна прекратить дальнейшие исследования, довольствуясь

предварительными данными, полученными благодаря небольшому числу наблюдений. Такие результаты могут носить лишь временный характер, и в ходе последующего анализа закономерности, выявленные на предварительных этапах, могут исчезнуть или стать диаметрально противоположными. Вот почему решение о преждевременном прекращении исследования нужно принимать с осторожностью и с учетом многих обстоятельств. Попытки комплексного учета промежуточных результатов клинических испытаний реализуются на основе ряда статистических подходов, однако следует понимать, что само по себе обсуждение проблемы модификации или прекращения испытания возможно лишь в случае выявления достоверной ассоциации. Статистический тест должен рассматриваться не как единственная основа для принятия решения, а скорее как сигнал к ревизии полученных промежуточных результатов. Нет единого мнения о том, какой статистический критерий должен рассматриваться в качестве такого сигнала. В качестве возможного подхода к оценке значимости полученных результатов можно использовать методику, основанную на последовательном анализе (Armitage P. [et al.], 2001). В тех случаях, когда предметом динамического мониторинга является изучение какого-либо количественного показателя с распределением, близким к нормальному, предлагается формула:

$$\delta = \frac{\theta}{\sqrt{\left(\frac{2\sigma^2}{m}\right)}}$$

где δ — характеристическая функция, отражающая степень достоверности и статистическую силу исследования при заданном числе пересмотров результатов; θ — среднее значение изменения анализируемого показателя в группе, подвергшейся интервенционному воздействию; σ — исходное среднеквадратическое отклонение изучаемого показателя; m — численность обследуемой и контрольной групп на каждой стадии.

При этом предполагается равное число наблюдений в группе, подвергшейся интервенционному вмешательству, и в контрольной группе наблюдений, а численность сопоставляемых групп практически не меняется на каждой стадии исследования. Значения характеристической функции δ для различных значений числа пересмотров результатов N при уровне статистической достоверности $2\alpha = 0,05$ и статистической силе теста $1 - \beta = 0,95$ представлены в табл. 3. В тех случаях, когда вычисленное значение δ превышает критический уровень, указанный в таблице, результаты можно считать достаточно надежными для прекращения дальнейших интервенционных испытаний.

**Критическое значение характеристической функции δ
в зависимости от числа пересмотров N**

Число пересмотров N	Характеристическая функция δ	Число пересмотров N	Характеристическая функция δ
2	2,66	7	1,51
3	2,22	8	1,42
4	1,95	9	1,34
5	1,76	10	1,28
6	1,62		

Расчет модифицируется в тех случаях, когда исследуемый эффект имеет не количественное, а качественное (категориальное) выражение. Для этого потребуется ряд предварительных преобразований. Если обозначить частоту распознаваемого состояния (колеблется в диапазоне от 0 до 1) в группе, подвергшейся интервенционному воздействию, и контрольной группе, соответственно, как λ_1 и λ_2 , то функция δ вычисляется следующим образом:

$$\delta = \theta \sqrt{\frac{m}{(\lambda_1^{-1}) + (\lambda_2^{-1})}},$$

а показатель θ , в свою очередь, рассчитывается на основании преобразования:

$$\theta = \ln\left(\frac{\lambda_1}{\lambda_2}\right).$$

Очевидно, показания к преждевременному прекращению интервенционных исследований не могут быть абсолютно идентичными во всех случаях, и оценки, основанные на статистических подходах, не должны рассматриваться как безусловные основания для принятия решения. Вот почему статистическая оценка должна быть проанализирована в контексте всех имеющихся доказательств ее значимости, в частности с учетом известных или предположительных биологических механизмов, а также результатов других известных исследований в конкретной области.

Факторы блокирования (*blocking factors*) отражают индивидуальные различия пациентов по определенным переменным в рамках общего для всех дизайна (например, специфическая симптоматика).

Учет таких факторов блокирования также именуется стратификацией. Их можно выделить в ходе проведения рандомизированного интервенционного исследования. Факторы блокирования включаются в анализ в качестве потенциальных модераторов или для сбалансирования экспериментальных групп по принципиально важным показателям, которые могут отразиться на результатах исследования. Процедура исследования должна быть организована таким образом, чтобы участники исследования были сгруппированы в соответствующие категории до того, как будет осуществлена их рандомизация на предмет отношения к экспериментальному воздействию. Например, такими факторами блокирования (стратификации) выступали пол пациента и выраженность исходного уровня дистресса (Sloane R. B., 1975). Исследователи предварительно распределили участников исследования по 4 ячейкам таблицы 2×2 (мужчины/женщины \times высокий/низкий уровень дистресса). Затем участников, вошедших в каждую такую ячейку, рандомизировали на предмет применения к ним одного из двух экспериментальных психотерапевтических подходов (поведенческая или психодинамическая терапия) или к нахождения в листе ожидания (контрольная группа).

Терминологически группа, в которой проводится изучаемое интервенционное воздействие, обозначается как **основная (экспериментальная, опытная, активная)** (Кельмансон И. А., 2002; Власов В. В., 2004; Покровский В. П., Брико Н. И., 2012; Rothman K. J. [et al.], 2008; Woodward M., 2014). Для обозначения альтернативной группы в экспериментальных исследованиях часто используются понятия **контрольная группа** и **группа сравнения**. Строго говоря, о контрольной группе следует говорить тогда, когда ее члены не подвергаются активному интервенционному воздействию, которое испытывают члены основной группы, и не получают какого-то иного активного воздействия. Термин «группа сравнения» предполагает, что ее члены получают какое-то альтернативное активное воздействие, отличное от того, которое получают представители основной группы. Однако часто используют обобщающий термин «контрольная группа», рассматривая в целом членов этой группы как альтернативу основной группы.

Контрольная группа, не получающая лечения (*no-treatment controls*), обеспечивает максимум контраста по отношению к исследуемому терапевтическому воздействию. Однако в ряде случаев использование в экспериментальных исследованиях такой контрольной группы сопряжено с рядом этических проблем, так как предполагает отказ от какого-либо лечения в отношении лиц, которые могут в нем нуждаться. **Контрольная группа, включающая пациентов из листа ожидания** (*wait-list controls*), может представлять некото-

рый компромисс, особенно в тех случаях, когда изучаемое лечение является краткосрочным и пациенты не страдают тяжелыми заболеваниями. В этой ситуации пациенты, включенные в лист ожидания на основе рандомизированного отбора, проходят то же первоначальное обследование, что и пациенты из основной группы, а затем относительно недолго ждут своей очереди до того момента, когда завершится лечение лиц, вошедших в основную группу.

Преимущества рандомизации по сравнению с иными принципами распределения испытуемых:

— При надлежащем соблюдении принципа рандомизации никто из исследователей, принимающих участие в исследовании, не будет заведомо осведомлен о том, из каких конкретных испытуемых будет состоять группа, подвергаемая изучаемому интервенционному воздействию. Как следствие, сводится к минимуму систематическая ошибка, связанная с тактическим выбором в отношении отдельного испытуемого. Такой ошибки не удастся избежать при использовании иных принципов распределения обследуемых по отношению к изучаемому интервенционному воздействию.

— Соблюдение принципа рандомизации обеспечивает сопоставимость анализируемых групп по важнейшим характеристикам обследуемых, а имеющиеся различия сводятся лишь к непосредственно анализируемой тактике интервенционного воздействия. Данное обстоятельство весьма важно, так как многие характеристики, отражающие стартовые условия, в которых находятся испытуемые, во многом могут предопределять успех интервенционного воздействия. Важно отметить, что соблюдение принципа рандомизации позволяет не только нивелировать влияние таких заранее учтенных факторов, но и при большом числе наблюдений практически сводит к нулю и возможное влияние иных неучтенных и зачастую непредсказуемых факторов, которые могли повлиять на результаты (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008).

Практические ограничения рандомизированного исследования:

— Рандомизированное отнесение пациентов к основной или контрольной группе не дает полной уверенности в том, что группы окажутся эквивалентными с самого начала или останутся таковыми по ходу исследования.

— Многие экспериментальные исследования по ходу их проведения теряют часть испытуемых, что связано с их отказом от участия в исследовании до окончания лечения и сбора информации о последующем тестировании. Такое уменьшение численности уменьшает эквивалентность основной и контрольной групп.

— Возможен и так называемый феномен перетекания между сопоставляемыми группами. Если пациенты, находящиеся в одной

больничной палате и составляющие основную группу, получают какое-то полезное воздействие (например, обучаются технике релаксации), они могут рассказывать об этом пациентам из другой больничной палаты, входящим в контрольную группу, обучая и их этой технике. В ходе проведения клинических испытаний описываются даже такие случаи, когда пациенты, получающие экспериментальные лекарственные средства, делятся этими лекарствами с пациентами из контрольной группы.

— Больничный персонал может не понять значения рандомизации в некоторых ситуациях, усматривая в этом нарушение этического принципа уважения автономии пациента, необходимости индивидуализированного подхода к проведению лечения.

— Рандомизированные исследования являются дорогостоящими и требуют больших временных затрат, поэтому используются в тех случаях, когда имеются предварительные свидетельства о пользе изучаемого экспериментального воздействия.

— Рандомизация по этическим соображениям не может быть использована для изучения влияния негативных воздействий (курения, потребления наркотиков, психологической травмы и т. п.).

— Рандомизированные исследования не принимают во внимание выбор пациента (Cook T. D., Campbell D. T., 1979; Rossi P. H. [et al.], 2004).

В ряде исследований после осуществления рандомизации и испытания авторы исключают из последующего анализа тех субъектов, которые плохо соблюдали условия протокола или по ряду признаков не удовлетворяли требованиям, предъявляемым к участникам. Такой подход следует признать тактически неверным. Исключение любого отобранного на основе принципа рандомизации пациента может привести к систематической ошибке, так как в любом исследовании абсолютно дисциплинированные пациенты составляют лишь часть наблюдений, а мотивы отказа от соблюдения условий протокола могут быть тесным образом связаны с вероятностью реализации изучаемого эффекта. Более того, идея интервенционного исследования предполагает поиск ответа на вопрос о том, в какой мере предложенная интервенционная программа сопряжена с положительным эффектом. Вот почему сама по себе рандомизация осуществляется, исходя из принципа подразделения испытуемых на группы, в которых предлагается или не предлагается к реализации анализируемая интервенционная программа, а не по принципу реализации или нереализации этой программы. Следовательно, искусственное сужение рамок анализируемой популяции неизбежно приведет к снижению статистической силы рандомизации, и истинные сравнения возможны только между сопоставляемыми группами в целом. После

осуществления рандомизации необходимо включать в последующий анализ всех испытуемых: «once randomized, always analyzed» (Hennekens С. Н. [et al.], 1987). В связи со сказанным становится очевидной необходимость обеспечения достаточно высокого уровня соблюдения участниками исследования условий протокола, сведения к минимуму потерь участников исследования в ходе динамического наблюдения за ними, а также сбора максимально полной информации. Если есть возможность, необходимо продолжать наблюдение даже за теми лицами, которые отказались от дальнейшего соблюдения условий исследования, что позволит в будущем осуществить сопоставление важнейших характеристик пациентов в зависимости от их склонности к сотрудничеству.

2.4. Открытое, слепое, двойное слепое исследование

Важной задачей интервенционных исследований является обеспечение правильной оценки предполагаемых последствий изучаемого воздействия. Для этого, прежде всего, соблюдается принцип, согласно которому информация в сопоставляемых группах должна быть представлена в одинаковом объеме, что позволит избежать систематической ошибки, связанной с более полным обследованием какой-то отдельной группы. Кроме того, необходимо достаточно продолжительное наблюдение за всеми участниками исследования. Чаще всего для окончательных результатов интервенционных исследований требуются месяцы и годы наблюдений, что особо актуально в случае хронических заболеваний, а также анализа данных, связанных с изучением продолжительности жизни. Иногда предполагаемый период наблюдения может быть коротким, например ограничиваться сроками пребывания обследуемого в стационаре, как это бывает в случае изучения возможных ближайших осложнений лекарственной терапии, и тогда легко обеспечить достаточную полноту информации в анализируемых группах. Очевидно, трудности, связанные с динамическим наблюдением за каждым обследуемым, прямо пропорциональны длительности предполагаемого наблюдения. Методы, направленные на обеспечение высокого процента сбора информации в ходе динамического наблюдения за участниками интервенционных исследований, принципиально совпадают с таковыми, используемыми в проспективных когортных исследованиях.

Вероятность ошибки в оценке результатов интервенционного воздействия зачастую связана с тем, что осведомленность исследователя о предполагаемом положительном или отрицательном эффекте может повлиять на его склонность к обнаружению такого эффекта. Это особо актуально в тех случаях, когда оценка эффекта имеет субъ-

активный характер. Так, если конечным пунктом интервенционного исследования является изучение влияния воздействия на смертность, можно ожидать минимального риска ошибки, связанной с субъективизмом, однако риск будет существенно выше, если результаты исследования предполагают оценку таких показателей, как жалобы больного, выраженность болевого синдрома, частота осложнений.

Минимизация указанной ошибки возможна в тех случаях, когда до полного окончания исследования его участники минимально осведомлены о степени подверженности каждого обследуемого изучаемому интервенционному воздействию, что получило наименование принципа **простого слепого исследования** (*single-blind design*). При **двойном слепом исследовании** (*double-blind design*) ни участники исследования, ни те исследователи, которые непосредственно отвечают за сбор информации о предполагаемом эффекте, не осведомлены о том, к какой группе в зависимости от интервенционного воздействия относится каждый отдельный испытуемый. В ряде случаев применяется и **тройное слепое исследование** (*triple-blind design*), при котором анализ результатов также осуществляется лицом, не осведомленным о том, к какой группе относится каждый отдельный участник исследования.

Возможность проведения слепого исследования зависит от того, в какой мере реально обеспечить внешнее сходство терапевтической программы и протокола последующего обследования всех участников. Вот почему во многих интервенционных исследованиях, прежде всего связанных с испытанием лекарственных средств, участники контрольной группы наблюдений получают **плацебо**. Проблема оценки конечного результата любой интервенционной программы связана с тем, что большинство участников склонно к некоторому оптимизму, независимо от истинного характера интервенционного вмешательства. Этот феномен получил название **плацебо-эффект**. Если в исследовании не используется плацебо-контроль, трудно с уверенностью сказать, являются ли субъективные данные, представляемые участниками, истинными, т. е. последствиями лечебного воздействия, отражают ли они обычное повышенное внимание к ним или же отражают уверенность пациентов в том, что лечение должно помочь. С другой стороны, при оценке побочных эффектов терапии плацебо-эффект может реализоваться со знаком «минус», когда пациенты, осведомленные о возможных побочных эффектах, будут склонны к активному поиску и обнаружению их у себя. Все указанные проблемы нивелируются при осуществлении двойного слепого исследования. Естественно, что такое исследование является более трудоемким. Кроме того, должна быть предусмотрена возможность немедленного информирования («прозрения») исследователя, на-

блюдающего за испытуемым, в отношении того, к какой группе относится пациент, если у последнего возникают какие-то осложнения, серьезные заболевания, неотложные состояния, при которых необходимы точные сведения о характере проводимой терапии.

В ряде случаев представляется затруднительным осуществление двойного слепого исследования; это, прежде всего, относится к тем ситуациям, когда интервенционное воздействие предполагает изменение образа жизни, диеты, проведение специфических лечебных мероприятий, психотерапевтических вмешательств. В указанных случаях возможно проведение простого слепого (*single-blind*) или открытого (*un-blind, open trial*) исследований. В случае простого слепого метода о характере осуществляемого интервенционного вмешательства у каждого пациента осведомлен только исследователь; в случае открытого метода информация доступна как исследователю, так и испытуемому. Подобные подходы технически более доступны, однако сопряжены с рядом проблем научного свойства. Так, испытуемые, не вошедшие в основную группу, могут быть разочарованы указанным обстоятельством, что приводит к большей вероятности их отказа от последующего участия в исследовании. Осведомленность о проводимой терапии может индуцировать пациентов в отношении их субъективных оценок положительных и отрицательных эффектов. Вот почему в случае неизбежности простых слепых или открытых исследований необходимо стремиться к оценке результатов на основе четких объективных критериев. Персонал, осуществляющий обследование, не должен располагать точными сведениями о том, кто из пациентов подвергался интервенционному воздействию.

2.5. Исследования в параллельных, связанных группах и перекрестный дизайн

Исследование с параллельным контролем — простейший вариант интервенционных исследований, который предполагает сравнение результатов вмешательства в двух группах — основной и контрольной. Основная группа получает изучаемое вмешательство, а контрольная — нет. Поскольку пациенты включаются в группы независимо друг от друга, такие группы называют независимыми. Полученные данные также называют независимыми (несвязанными).

Любые систематические различия между участниками сопоставляемых групп могут повлиять на результаты, это является основным ограничением исследования в параллельных группах. Хотя рандомизация и позволяет минимизировать такие систематические различия, их присутствие нельзя полностью исключить, особенно по таким важным параметрам, как возраст, пол, масса тела, рост обследуемых пациентов.

Эти факторы могут иметь прогностическое значение и сказаться на результатах исследования, особенно при малой численности выборки.

Исследование в связанных выборках (подобранных парах, копиях-парах, *matched pairs*) — вариант исследования, при котором каждому пациенту в одной группе соответствует парный ему пациент в другой. Парой может быть пациент, подобранный, например, по полу и равному возрасту. Когда подобрана пара, проводится рандомизация, которая должна определить, кто из членов пары получит изучаемое интервенционное воздействие, а кто станет представителем контрольной группы. Пациенту основной группы может соответствовать не один, а два пациента и более в контрольной. Такая организация исследования призвана преодолеть систематические различия между испытуемыми, которые могут иметь прогностическое значение и повлиять на результаты. Неблагоприятным последствием такого подбора может быть избыточный подбор (*overmatching*), когда пары подбирают по признакам, непосредственно связанным с изучаемыми особенностями болезни. В таком случае изучаемые особенности могут быть замаскированы. Например, если контрольные пациенты были подобраны по равной массе тела, то выявить влияние питания на исход болезни будет затруднительно (Власов В. В., 2004). Принципиально важно, что подбором пар можно устранить влияние лишь известных (предполагаемых) побочных факторов, но нельзя с уверенностью исключить действия других неучтенных факторов. Поэтому одного лишь подбора пар недостаточно для обеспечения валидности исследования, и интервенционное воздействие внутри отдельных пар следует определять с использованием принципа рандомизации.

Перекрестный (гибридный) дизайн исследования (*cross-over study*) состоит в том, что изучаемое воздействие попеременно в различные периоды назначается всем участникам исследования (Senn S., 2002; Woodward M., 2014). Дизайн призван устранить ограничения, указанные для предыдущего дизайна. Наиболее простым вариантом такого дизайна исследования является вариант «два периода, два лечения» (или 2×2). В случае дизайна 2×2 пациенты распределяются на группы А и В на основании рандомизированного подхода. Пациенты в группе А получают лечение 1 в течение определенного периода времени, а затем получают лечение 2. Пациенты в группе В получают лечение в противоположной последовательности (сначала 2, затем 1) (рис. 28). Как правило, продолжительность курсов лечения 1 и 2 одинакова (обычно составляет несколько недель). Простой слепой дизайн может быть реализован, если применить технику дублирования пустышкой (*double dummy*): каждый участник всегда в ходе исследования получает комбинацию двух лечебных схем: реального

активного воздействия, проводимого ему в настоящее время, и плацебо-воздействия, внешне не отличимого от того альтернативного лечения, которое пациенту в данный момент не проводится.

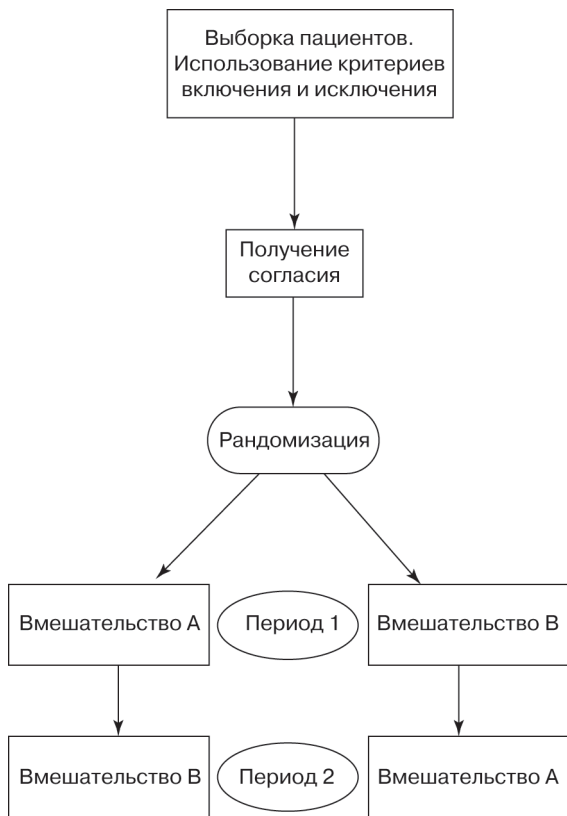


Рис. 28. Схема перекрестного дизайна интервенционного исследования

Достоинства перекрестного дизайна

— Эффект воздействия 1 может быть сопоставлен с эффектом воздействия 2 у каждого обследованного субъекта. Такие различия, зафиксированные у каждого испытуемого (*within-subject differences*), затем суммируются, что позволяет оценить общую эффективность интервенционного воздействия.

— Каждый испытуемый выступает в качестве контрольного наблюдения для самого себя, что устраняет проблему различий между участниками сопоставляемых групп.

— Можно уменьшить число обследуемых пациентов, так как основная и контрольная группы — это одни и те же лица.

Недостатки перекрестного дизайна

— Перекрестный дизайн может использоваться только для изучения состояний с длительным течением, для которых лечение может сопровождаться лишь краткосрочным эффектом (о полном излечении вообще не может быть речи), в противном случае не будет возможности оценки второго альтернативного воздействия.

— Каждый обследуемый субъект участвует в исследовании довольно долго (до двух раз дольше, чем при исследовании в параллельных группах). Такая ситуация повышает риск отказа от участия в исследовании, нарушения протокола исследования испытуемыми.

— После завершения первого изучаемого воздействия у пациентов в течение определенного времени могут сохраняться его остаточные эффекты, проявляющиеся уже в ходе испытания второго воздействия (*carry-over effect*). Если воздействие 1 является активным, то после его завершения у пациентов могут сохраняться положительные последствия проведенного лечения, которые проявятся уже при переходе к назначению им воздействия 2 — плацебо, что может исказить результаты. С целью избежать этого остаточного эффекта многие исследования перекрестного дизайна предусматривают «период вымывания» (*washout period*) между двумя этапами исследования. Иногда в «периоде вымывания» пациентам назначают плацебо, и этот период не учитывается при последующем анализе результатов. Остаточные эффекты лечения могут взаимодействовать и с естественным течением заболевания, т. е. назначаемое лечение может по-разному проявляться в отдельные периоды заболевания.

— Статистический анализ исследований перекрестного дизайна более сложен, чем в случае исследований параллельных групп. Такой анализ обычно предполагает использование парных критериев (парный критерий t Стьюдента или его аналог — непараметрический критерий Уилкоксона). Детальное описание статистических подходов к анализу перекрестных интервенционных исследований приводится в специальных руководствах (Senn S., 2002; Woodward M., 2014).

2.6. Экспериментальный дизайн с одним испытуемым

Экспериментальный дизайн с одним испытуемым используется в тех случаях, когда требуется проведение интервенционного исследования на одном пациенте. Субъект исследования является контролем для самого себя. Данный дизайн также иногда обозначается как

Низ 1 (N-of-1), так как позволяет получить N наблюдений на основе одного обследуемого пациента.

Экспериментальный дизайн с одним испытуемым предполагает, что проводятся повторные измерения и мониторинг изменений анализируемых показателей у испытуемого. Прежде всего, выбирается показатель или показатели, которые следует измерять у пациента. Показатели должны быть такими, чтобы их можно было часто повторно исследовать у данного пациента. Если речь идет о психометрических методиках, то они должны быть относительно краткими и минимально реактивными. Чаще всего применяются оценки, даваемые наблюдателями (например, оценка выраженности симптомов аутоагрессии) или самим испытуемым на основании самонаблюдения. Далее следует определиться с оптимальной частотой проведения измерений. Как правило, такие измерения проводят ежедневно. Однако в зависимости от задач исследования замеры можно проводить чаще (например, ежечасно) или реже (например, еженедельно). При любом варианте организации исследования начинают с серии базовых измерений, направленных на выявление исходных значений параметров, интересующих исследователя, до начала интервенционного воздействия. Базовые измерения проводятся до достижения стабильных показателей (предпочтительно от 10 до 20 наблюдений). После этого начинается фаза экспериментального лечения. Можно выделить несколько наиболее распространенных дизайнов:

— **Дизайн АВ** (*AB design*) — простейшая форма эксперимента с одним обследуемым. Проводится серия базовых измерений (фаза А), после чего осуществляется интервенционное воздействие (фаза В), и на этой фазе также осуществляется серия измерений (рис. 29). На

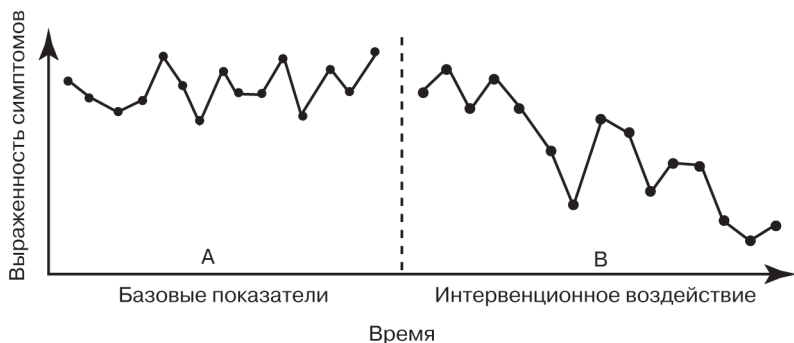


Рис. 29. Дизайн экспериментального исследования с одним испытуемым по типу АВ (Barker C. [et al.], 2002)

основе дизайна можно, например, исследовать влияние позитивного стиля воспитания на проявления аффективных вспышек у ребенка 6 лет. Родителям предлагают фиксировать число и выраженность таких аффективных вспышек у ребенка ежедневно в течение 2 нед. (фаза А). Затем в ходе фазы В (интервенционное воздействие) родителям может быть предложен новый способ реагирования на аффективные вспышки у ребенка. Если интервенционное воздействие окажется эффективным, можно ожидать снижения частоты и выраженности аффективных вспышек.

Недостатком дизайна АВ является то, что при отсутствии дополнительной информации такое исследование не позволяет с уверенностью говорить о существовании причинно-следственной связи между воздействием и эффектом. На него в значительной степени распространяются те критические соображения, которые были приведены в отношении квазиэкспериментального дизайна по типу «одна группа, предварительное тестирование, последующее тестирование». Так, в приведенном примере на момент начала интервенционного вмешательства у ребенка мог появиться новый друг, что привело к снижению частоты аффективных вспышек.

— **Реверсивный дизайн (АВАВ) (*reversal design*)** — дизайн АВ с повторением. Такой дизайн довольно часто используется в исследованиях, связанных с модификацией поведения (рис. 30).

Пример

В примере с изучением аффективных вспышек у ребенка дизайн АВАВ можно было бы реализовать следующим образом: после выявления эффективности интервенционных воздействий нужно было бы вернуться к исходной базовой стадии, а затем вновь осуществить интервенционное воздейс-

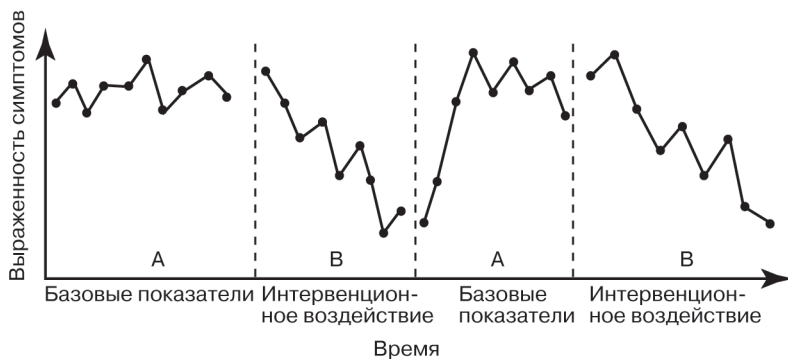


Рис. 30. Дизайн экспериментального исследования с одним испытуемым по типу АВАВ (Barker С. [et al.], 2002)

твие. Вмешательство, связанное с воспитанием ребенка, окажется эффективным, если при прекращении такого вмешательства у ребенка учащаются аффективные вспышки.

Обоснованием такого подхода является предположение, что изменение поведения после реверсий свидетельствует о влиянии экспериментального вмешательства. Существуют и более сложные модификации реверсивных дизайнов, например АВАСАВ, в котором после базовой фазы А используется второе вмешательство С.

Недостатки реверсивного дизайна

— Может иметь место остаточный эффект. У пациентов автоматически не возникает обострений заболевания сразу после окончания лечения. Это особенно актуально в случаях изучения изменения поведения пациентов на фоне психодинамической и когнитивной терапии, поскольку такие психотерапевтические вмешательства, если они эффективны, приводят к необратимым изменениям стиля мышления пациента.

— Даже если интервенционное воздействие является обратимым, возникают серьезные этические проблемы, связанные с прекращением потенциально эффективного лечения на втором и последующем переходах к базисным фазам исследования. Таким образом, данный дизайн создает конфликт клинических и научных интересов.

— Чередующиеся назначения и отмены изучаемого лечения может вызвать нежелательный психологический эффект (пациент может утратить доверие к специалисту, может быть спровоцировано нежелательное поведение).

— **Дизайн с несколькими базовыми показателями** (*multiple baseline design*) используется при исследовании нескольких (предположительно независимых) поведенческих характеристик или какого-то одного стиля поведения, проявляющегося в различных ситуациях (рис. 31). Например, можно исследовать влияние интервенционных воздействий одновременно на проявления аффективных вспышек, энуреза и страхов у ребенка или на проявления агрессии в домашних условиях, в школе и на игровой площадке. При этом осуществляются интервенционные воздействия, направленные на каждое изучаемое нарушение поведения, и оценивается их эффект в отношении каждого отдельно изучаемого вида нарушения поведения (см. рис. 31). Дизайн предполагает, что каждое отдельное интервенционное воздействие может оказаться эффективным в отношении лишь какого-то одного стиля поведения или одной конкретной ситуации, в которой этот стиль поведения проявляется. Его нельзя признать удовлетворительным в тех случаях, когда изучаемое интервенционное воздействие направлено на изменение каких-то общих показателей у пациента.

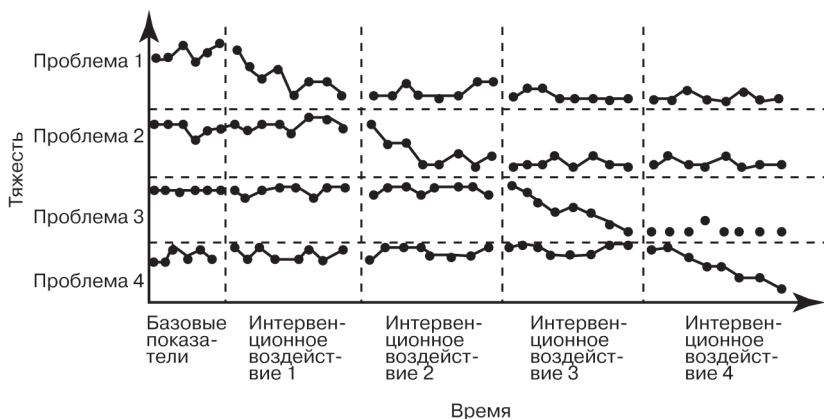


Рис. 31. Дизайн экспериментального исследования с одним испытуемым и несколькими базовыми показателями (Barker C. [et al.], 2002)

— **Дизайн с переменным критерием** (*changing-criterion design*) используется для демонстрации того, что интервенционное воздействие может прогрессивно снижать выраженность нарушения, т. е. по мере ужесточения критерия эффективности (рис. 32). Дизайн особенно полезен при работе с пациентами, страдающими аддиктивными расстройствами. Например, при разработке программы по отказу от курения на первом этапе в качестве критерия эффективности интервенционного воздействия может быть использовано снижение числа выкуриваемых сигарет до 15 в день, на следующем этапе — до 10 в день (Barker C. [et al.], 2002; Nikles J., Mitchell G., 2015).

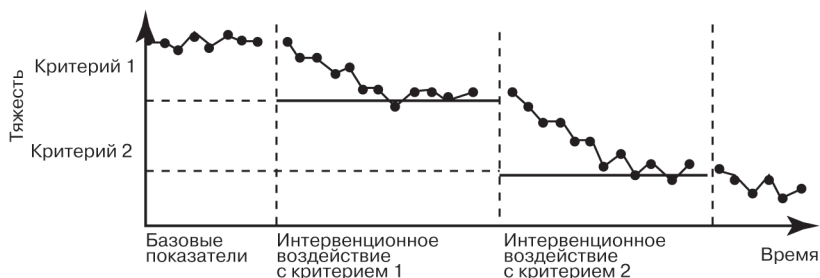


Рис. 32. Дизайн экспериментального исследования с одним испытуемым и переменным критерием (Barker C. [et al.], 2002)

Анализ исследований с одним испытуемым обычно предполагает графическое отображение результатов. Предложены некоторые статистические подходы к анализу данных, полученных в ходе подобных исследований (Morley S., Adams M., 1989). Одним из таких подходов может быть рассмотренная ранее методика выявления точек перелома (join points) при анализе тренда (см. Корреляционные исследования). Анализ показателей, приведенных для иллюстрации дизайна АВ (см. рис. 29), на предмет наличия точек перелома при помощи программы JoinPoint (рис. 33) выявил наличие одной точки перелома, соответствующей 16-му наблюдению (начиная с 15-го наблюдения оказывалось экспериментальное интервенционное воздействие). С этого момента выявлено наличие статистически достоверного тренда к снижению показателей ($P < 0,0001$). В первой фазе наблюдений (с 1-го по 15-е) отсутствовала статистически достоверная динамика показателей ($P = 0,33$).

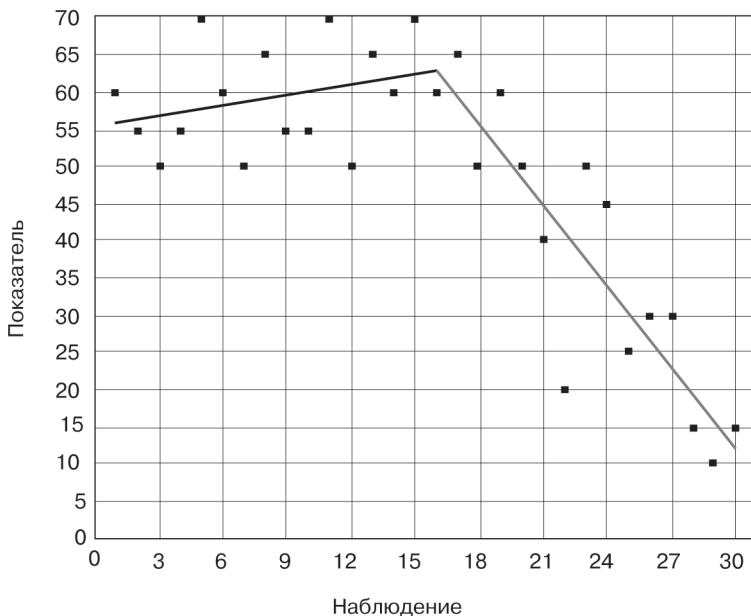


Рис. 33. Выявление «точки перелома» при экспериментальном исследовании с одним испытуемым и дизайном АВ

Вопросы для самоконтроля

1. Что такое дизайн исследования?
2. Какие существуют варианты описательных исследований?
3. В чем суть исследования популяционных срезов? Какие подходы используются при формировании выборки в ходе исследований популяционных срезов?
4. Что понимают под корреляционными исследованиями?
5. Что такое аналитические исследования? В чем принципиальные отличия исследований по принципу «случай — контроль» от когортных исследований?
6. Что такое интервенционные исследования? В чем различия квазиэкспериментального и рандомизированного дизайнов?
7. Что такое открытое, слепое, двойное слепое исследование?
8. Что означают исследования в параллельных группах, исследования в связанных группах, перекрестный дизайн?
9. В чем суть экспериментального дизайна с одним испытуемым?

Глава 3

ПРОЦЕСС ИЗМЕРЕНИЯ

Признаки, изучаемые в ходе клинико-психологических исследований, можно условно сгруппировать в следующие важнейшие категории:

— когнитивные (мысли, аттитюды, надежды, ожидания, атрибуции, память, мышление и т. п.);

— аффективные (чувства, эмоции, настроение, телесные ощущения и т. п.);

— поведенческие (действия, навыки, умения, речь и т. п.);

— биологические: анатомо-физиологические (частота сердечных сокращений, артериальное давление, иммунологические показатели и т.п.);

— социальные: острые или хронические стрессоры, социальная поддержка, семейное функционирование, работа и т. п. (Barker С. [et al.], 2002).

Когда используется понятие «измерение», то в общем смысле предполагается процесс нахождения индикаторов для тех или иных психологических концептов, т. е. понятий. Как правило, все психологические концепты имеют многочисленные индикаторы. Например, концепт «фобия» предполагает изучение признаков избегания, указаний на наличие страхов, выявление физиологических показателей, отражающих активацию симпатической нервной системы на фоне предъявления соответствующих стимулов. Концепт «боль» предполагает оценку самим испытуемым интенсивности боли, выявление болевого поведения, клинические оценки. Концепт «страх смерти» предполагает выявление указаний на эти страхи по данным опросников.

Психологическое понятие (концепт), лежащее в основе изучаемого показателя, часто именуется конструктом. **Конструкт** — гипотетический фактор, который невозможно наблюдать непосредственно, его существование можно предположить на основании определенно-го поведения и определенных обстоятельств (Гудвин Д., 2004).

Мера конструкта — признак, индикатор, позволяющий промерить данный конструкт. Приведенная терминология чаще всего используется в количественных методах измерений, но применима и в качественных методах, хотя в последнем случае чаще говорят не о конструкте, а о **феномене**. **Операционализация** (*operationalization*) (Barker С. [et al.], 2002) — процесс перехода от конструк-

та к ассоциированной с ним мере. Для операционализации того или иного конструкта часто имеется несколько различных путей. Выбор подхода к измерению конструкта зависит от вопросов, стоящих перед исследователем, от определенных теоретических предпосылок и доступных ресурсов. Кроме того, следует понимать, что ни одно отдельно взятое измерение не может всесторонне охарактеризовать изучаемый конструкт. Наконец, нужно учесть, что связь между измерением и конструктом, как правило, не носит линейного характера: на измерение конструкта влияет то, как мы его определяем и как себе представляем. Для упрощения процесса операционализации конструкту может быть дано **операционное определение** (*operational definition*), при помощи которого его будет легче измерить. Так, конструкт «эмпатия» можно определить следующим образом: «Быть в состоянии эмпатии означает воспринимать внутренний мир другого точно, с сохранением эмоциональных и смысловых оттенков. Как будто становишься этим другим, но без потери ощущения “как будто”. Так, ощущаешь радость или боль другого, как он их ощущает, и воспринимаешь их причины, как он их воспринимает. Но обязательно должен оставаться оттенок “как будто”: как будто это я радуюсь или огорчаюсь. Если этот оттенок исчезает, то возникает состояние идентификации» (Роджерс К., 1984). Для процесса психотерапии операционное определение может быть следующим: эмпатия — то, насколько точно психотерапевт реагирует на чувства, выраженные клиентом. Признаки последнего могут быть выявлены методом наблюдения: путем экспертных оценок зарегистрированных видеозаписей общения психотерапевта и клиента.

Гарвардский физик П. Бриджмэн отмечал, что научная терминология должна быть абсолютно объективной и точной, а все понятия должны быть определены в терминах выполняемых операций. В 1920-х гг. он впервые ввел термин **«операционизм»**. Философская доктрина операционализма (операционизма) рассматривает научные концепции полностью с точки зрения рабочих (т. е. оперантных) процедур, с которыми они связаны. Примером операционализма служит следующее утверждение: «Интеллект — это то, что определяют тесты на интеллект». Несмотря на проблематичность строгого применения операциональных определений, теория операционизма сыграла свою роль в психологии и научила исследователей четко определять используемые термины (Гудвин Д., 2004).

В дальнейшем на смену концепции операционализма пришла **стратегия критического реализма — конвергентных (сходящихся) операций**, в соответствии с которой для оценки одного конструкта следует использовать многочисленные индикаторы. В клиническом контексте это означает, что исследователи часто не

могут адекватно изучить многие важные конструкты при помощи традиционно используемых подходов. Например, социальные навыки могут быть операционализированы такими индикаторами, как хороший контакт «глаза в глаза», улыбка и т. п., однако использование только этих характеристик не дает исчерпывающей информации об изучаемом конструкте; наоборот, такие характеристики могут быть типичны для людей, которые ведут себя как роботы. Вот почему с позиций критического реализма нужно признать, что многие психологические конструкты могут быть охвачены измерениями лишь отчасти. Из этого следует, что если ряд исследований, использующих несколько отличные операциональные определения и процедуры, приходят к одинаковым выводам, то есть больше оснований полагать, что конструкт изучен достаточно полно.

Оценки, полученные в ходе измерений, могут иметь как количественный, так и качественный характер. По определению, количественные оценки используют численные показатели.

Достоинства количественных оценок:

— Использование численных показателей обеспечивает большую точность измерений. Существует разработанная теория оценки надежности и валидности методов, использующих количественные оценки, что позволяет исследователю определить, в какой степени он может доверять полученным показателям.

— Имеются хорошо отработанные статистические приемы анализа количественных данных. Эти данные могут быть легко суммированы и представлены в доступной форме.

— Количественные оценки можно сопоставлять.

— Количественные оценки хорошо удовлетворяют гипотетически дедуктивному подходу к организации исследования. Гипотетическая связь между изучаемыми переменными может быть представлена при помощи математической модели, и метод статистической инференции допустимо использовать для оценки того, насколько хорошо фактические данные согласуются с предсказанной моделью.

— Допустимо использование статистических подходов для обоснования возможности обобщения полученных результатов, полученных в выборке из популяции, в отношении всей популяции (Barker C. [et al.], 2002).

3.1. Методы и методики измерения

Метод — подход (прием, средство) теоретического исследования или практического воплощения явления (процесса).

Методика — форма реализации метода, совокупность приемов и операций (их последовательность и взаимосвязь), процедура или

набор процедур для достижения определенной цели. В большинстве случаев для изучения сходных сторон объекта можно использовать различные методики, что обеспечивает взаимную верификацию данных, получаемых различными методиками. Этап выбора методик тесно связан с конкретизацией исследовательских гипотез. Поскольку психические явления непосредственно не наблюдаемы и не измеряемы (во всяком случае, изолированно от других явлений, присущих человеку), то в эмпирическом исследовании они могут быть обнаружены только через соответствующие каждому явлению, свойству индикаторы (признаки, референты) (Голев С. В., Голева О. С., 2010а).

Исходя из концепции критического реализма, при составлении набора методик целесообразно руководствоваться **принципом дополнительности**. Данные качественного и количественного характера, субъективные и объективные показатели должны дополнять друг друга. Важно, чтобы результаты, полученные у отдельных испытуемых, были сопоставимы.

Если в наборе есть методики, позволяющие получить качественные и количественные оценки, это облегчает интерпретацию. Интерпретировать результаты, полученные методиками, дающими информацию лишь качественного характера (к таковым относятся многие проективные методики), нередко оказывается затруднительным.

3.2. Основные группы методов

К основным группам методов психологической диагностики можно отнести следующие:

- методы измерения функциональных показателей, использования психофизиологических индикаторов психических явлений;
- методы наблюдения и самонаблюдения;
- праксиметрические методы, основанные на анализе продуктов деятельности;
- субъективно-оценочные, основанные на анализе ответов испытуемых на вопросы опросников, анализе выборов альтернативных суждений, выборов градаций шкал и т. п.;
- проективные, связанные с анализом особенностей интерпретации внешне неопределенного материала, становящегося объектом проекции;
- ассоциативные, основанные на анализе ассоциативных ответов и выборов испытуемых (Волков Б. С. [et al.], 2005; Голев С. В., Голева О. С., 2010а).

Важнейшие достоинства и недостатки указанных групп методов суммированы в табл. 4.

Достоинства и недостатки важнейших диагностических методов
(Голев С. В., Голева О. С., 2010а)

Достоинства	Недостатки
<p>— Психологические данные, полученные для разных индивидуумов в одинаковой ситуации или для одного индивидуума в различных ситуациях, поддаются объективному сравнению</p>	<p><i>Психологические индикаторы</i></p> <p>— Часто методы дают обобщенную информацию о состоянии организма обследуемых лиц;</p> <p>— методы чувствительны к многочисленным посторонним воздействиям, что способствует появлению артефактов;</p> <p>— такие методы трудно использовать при одновременном обследовании сразу нескольких испытуемых;</p> <p>— методы трудоемки, нуждаются в сложном, квалифицированном анализе, особенно в тех случаях, когда необходимо провести длительное наблюдение за несколькими обследуемыми на протяжении большого отрезка времени</p>
<p>— Обеспечивается широка охвата проявлений психики;</p> <p>— больше потенциальных возможностей опознать влияние стимуляции и диссимуляции</p>	<p><i>Методы наблюдения и самонаблюдения</i></p> <p>— Параметры психических явлений описываются опосредованно;</p> <p>— сложная система критериев оценки и алгоритма интерпретации, оставляющая место проявлениям субъективизма</p>
<p>Могут давать достаточно объективные и надежные данные</p>	<p><i>Практические методы</i></p> <p>— Многие важнейшие параметры состояния не находят адекватного отражения в результатах деятельности и могут получить инвертированное отображение;</p> <p>— точность измерения многих параметров мала</p>

<p>— Возможность охватить многие аспекты внутренней картины психического процесса или состояния;</p> <p>— доступна ранняя диагностика изменений психических явлений;</p> <p>— имеется большой перечень ситуаций, доступных для обследования отдельного человека и целых групп;</p> <p>— очевидна относительная простота обработки результатов обследования;</p> <p>— возможна однозначная интерпретация результатов</p>	<p><i>Субъективно-оценочные методы</i></p> <p>— Описывается описание субъективного восприятия и отношения;</p> <p>— описание основывается на личностных смыслах;</p> <p>— нередко испытуемые не умеют правильно оценить и ответ, бывают неискренними в ответах</p>
<p>Методы имеют ценность для формирования полной клинической картины</p>	<p><i>Проективные методы</i></p> <p>— Методы позволяют дать лишь общую характеристику личности или несут информацию об отдельных переживаниях, желаниях, представлениях индивида;</p> <p>— данные этих методов эффективны лишь при учете результатов других методов</p>
<p>То же, что у проективных методов</p>	<p><i>Ассоциативные методы</i></p> <p>То же, что у проективных методов</p>

3.2.1. Опросники

Опросники — стандартизованный измерительный метод, иначе — стандартизованный самоотчет (Голев С. В., Голева О. С., 2010а). Их использование является весьма распространенным субъективно-оценочным методом исследования. В зависимости от содержания (или направленности) вопросов, которые входят в опросники, выделяют:

— *вопросы о личности респондента*, касающиеся его пола, возраста, образования, профессии, семейного положения и т. д., позволяющие в дальнейшем обрабатывать материал анкетирования в пределах той или иной подгруппы людей, при необходимости сопоставляя сходную информацию из различных подгрупп;

— *вопросы о фактах сознания*, предназначенных для выявления мнений, мотивов, ожиданий, планов, оценочных суждений;

— *вопросы о фактах поведения*, выявляющих реальные поступки и результаты деятельности людей.

В зависимости от формы ответа вопросы подразделяются на закрытые, полузакрытые и открытые.

Закрытый вопрос содержит полный набор возможных ответов, респондент должен выбрать один из них. Количество вариантов, которые можно выбрать (один или несколько), обычно оговаривается инструкцией (Волков Б. С. [et al.], 2005; Голев С. В., Голева О. С., 2010а; Barker С. [et al.], 2002; Newby P., 2014).

Существуют следующие способы предъявления вариантов ответов закрытого вопроса:

— дихотомическая форма, предлагающая противоположные, взаимоисключающие ответы («да» — «нет», «верно» — «неверно», «согласен» — «не согласен» и т. п.);

— поливариантная форма, предусматривающая «меню ответов», где можно выбрать несколько из них;

— форма шкалы, предназначенная для выражения интенсивности отношения, переживания, впечатления и т. п.; снабжение опросников вспомогательными шкалами позволяет получать количественные оценки;

— табличная форма: вопросы и возможные варианты ответов представлены в таблице.

Закрытые вопросы позволяют эффективно сопоставлять результаты респондентов и легче поддаются статистической обработке. Однако в закрытых вопросах часто отсутствует полнота выражений индивидуальных мнений или оценок, что может вызвать у испытуемого затруднение при поиске ответа на вопрос.

Полузакрытый вопрос используют, если составитель опросника не осведомлен обо всех возможных вариантах ответов или же намерен

более точно и полно выяснить индивидуальные точки зрения обследуемых лиц. Кроме перечня готовых ответов такой вопрос содержит графу «другие ответы» и определенное количество пустых строк.

Открытый вопрос предполагает, что ответ на него целиком и полностью сформулирует сам респондент. При наличии открытых вопросов опросники становятся более гибкими, так как это позволяет респонденту максимально полно и всесторонне выразить свое мнение, но труднее сопоставить ответы отдельных респондентов.

Шкала — это упорядоченная совокупность значений величины, которая служит основой для ее измерения.

При формировании рейтинговых шкал респондент должен дать цифровую оценку определенному суждению. Существуют многочисленные подходы к формированию подобных шкал. Наибольшую популярность завоевала **шкала Лайкерта** (*Likert scale*) (Nunnally J. C., Bernstein I. H., 1994). При конструировании подобной шкалы, прежде всего, требуется решить вопрос о том, сколько делений должно быть представлено на шкале. Словесные шкалы, используемые для предъявления закрытых вопросов, представляют собой набор суждений (не менее двух) о степени выраженности признака:

— *Бинарная, или дихотомная, шкала (binary or dichotomous scale)*. Суждения могут быть представлены в форме полюсов шкалы (свернутая словесная шкала), например «сильный — слабый», «быстрый — медленный», «активный — пассивный».

— *Шкала с множественным выбором (multiple choice)*. Суждения могут быть представлены в форме упорядоченного набора словесных обозначений полюсов и градаций (ступеней) шкалы (развернутая словесная шкала). Развернутая словесная шкала предполагает словесное выражение широкого спектра степеней проявления или значений одного признака.

В некоторых случаях использование определенного числа предлагаемых ступеней логически обосновано: возможны ситуации, когда ответ на вопрос может быть лишь «да» или «нет». Однако значительно чаще число предлагаемых вариантов ответов в рамках шкалы определяется исследователем, которому следует учитывать при своем выборе ряд обстоятельств (Barker C. [et al.], 2002):

— надежность возрастает при увеличении ступеней шкалы (Nunnally J. C., Bernstein I. H., 1994), однако нужно помнить, что увеличение ступеней более 5 редко сопровождается существенным возрастанием надежности методики, а для многих респондентов затруднительно сделать свой выбор в случае, когда предлагается более 7 ответов;

— большинство людей стремится избегать крайних точек шкалы (центральная тенденция) (Barker C. [et al.], 2002): на практике

лучше иметь минимум 5 ступеней шкалы, так как при наличии 3 или 4 с большей вероятностью можно ожидать появления «серединных» ответов.

— *Визуально-аналоговая шкала (visual-analog scale)*. Респонденту предлагают на линии (например, длиной 10 см) отметить точку, соответствующую степени выраженности у него того или иного признака. Такой подход, в частности, используют при оценке степени выраженности боли (рис. 34).

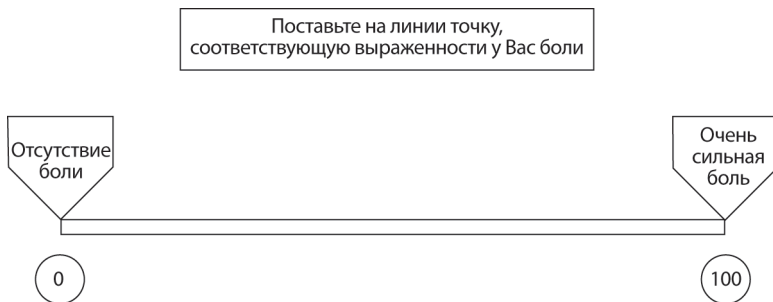


Рис. 34. Пример визуально-аналоговой шкалы

Шкалы могут быть однополярными и биполярными. *Однополярная шкала* ориентирована лишь на изучение одного конструкта и предлагает оценить степень выраженности признака (рис. 35). Так, шкала, направленная на оценку степени выраженности боли, может варьировать в диапазоне от «отсутствует боль» до «нетерпимая боль».

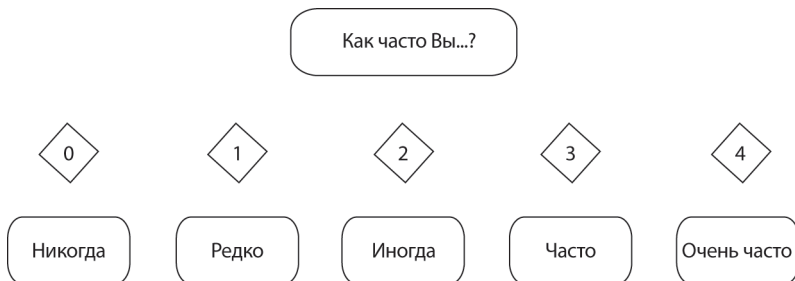


Рис. 35. Пример однополярной шкалы Лайкерта

— *Биполярная шкала* имеет противоположные дескрипторы на каждом конце шкалы (например, «активный» на одном конце, «пассивный» — на другом) (рис. 36). Биполярные шкалы могут иметь среднюю точку, представляющую некоторый нейтральный выбор («затрудняюсь ответить», «ни да ни нет»). Такой средней точки может и не быть, иными словами, шкала может иметь как нечетное, так и четное число ступеней. Некоторыми исследователями высказываются доводы против формирования такой средней точки, что связывается с феноменом форсированного выбора.

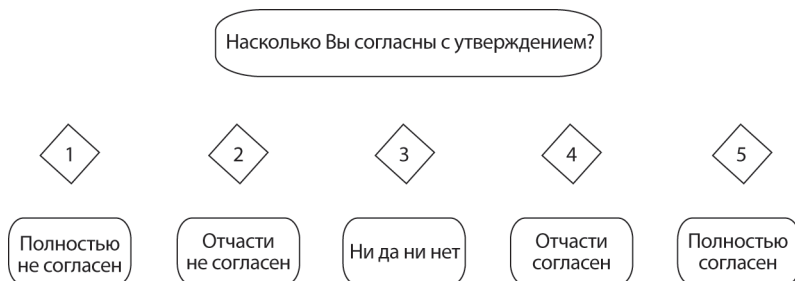


Рис. 36. Пример биполярной шкалы Лайкерта

Форсированный выбор (*forced choice*) (Barker С. [et al.], 2002) состоит в том, что у каждого респондента имеется какое-то определенное мнение, которое он может сформулировать, если его к этому слегка «подтолкнуть». Форсированный выбор облегчает анализ данных, так как респонденты условно делятся на тех, кто дает утвердительные или отрицательные ответы. Однако, если вопрос сформулирован удачно, при отсутствии средней точки и нейтрального варианта ответа приводит к тому, что нейтральные ответы могут реально отражать мнение части респондентов, и нет оснований лишать их возможности высказать такую точку зрения.

Анкеровка (*anchoring*) — создание своеобразных опорных пунктов — правильное обозначение словами и соответствующими им баллами тех точек, которые приведены на шкале опросника. Процедура вызывает большой практический интерес. Ее основу составляют некоторые допущения:

— шкала подчиняется закономерностям интервальной переменной: каждый шаг на такой шкале имеет равную величину (например, дистанция между градациями признака «вовсе нет» и «слегка» равна дистанции между градациями признака «очень» и «крайне»);

— все респонденты одинаково понимают формулировки, которые используются для обозначения градаций признака; например, на визуально-аналоговых шкалах исследователи часто анкеруют конечные точки шкалы и используют семантические дифференциалы (приводят парные противоположные по значению прилагательные: «хороший — плохой», «легкий — тяжелый» и т. п.).

Словесные суждения не всегда однозначно толкуются даже членами группы, сходными по полу, возрасту, социально-экономическому статусу и образованию. Эта неопределенность понимания и интерпретации суждений затрудняет их применение в шкалах в качестве точного психометрического инструмента.

Способы компенсации недостатков словесных шкал

— Необходимо согласовывать суждения с особенностями объекта. Например, оценивая память, лучше пользоваться понятиями «хорошая — плохая», чем «сильная — слабая», «высокая развитость — низкая развитость».

— Стараться избегать неточности, расплывчатости, двусмысленности в суждениях, понятиях, обозначениях полюсов шкал и промежуточных градаций. Например: «сильный», «средний», «слабый интерес» лучше, чем «очень сильный — очень слабый интерес», «значительный — незначительный интерес».

— Словесные обозначения не должны пересекаться, чтобы в одном понятии не содержалась часть другого, например: «добросовестно», «прилежно», «основательно»; «довольно сильно», «чересчур умный».

— Словесные обозначения, по возможности, не должны содержать оценок, например: «очень известный» — «малоизвестный», «очень толстый» — «очень стройный».

— Словесные обозначения должны указывать на один признак, а не на ряд признаков объекта. Например, неудачной является шкала с полюсами: желание учиться — нежелание учиться. Лучше разбить ее на 2 или 3 однородные по смыслу шкалы: интерес к учебе, трудолюбие в освоении учебных предметов, успешность в учебе (Голев С. В., Голева О. С., 2010а).

Способ формулирования. В зависимости от способа формулирования вопросы могут быть прямыми и косвенными. *Прямой вопрос* направлен непосредственно на респондента и нацелен на открытое получение информации от него. Предполагается, что на такой вопрос будет получен прямой и честный ответ. Однако там, где требуется выразить достаточно критическое отношение к себе и к другим, многие склонны ограничиваться социально одобряемыми ответами подчас в ущерб искренности. В таких случаях составляют *косвенный вопрос*, который обычно связан с использованием какой-либо вообразяе-

мой ситуации, маскирующей критический потенциал передаваемой информации.

Функции. Вопросы делятся на группы в зависимости от функции:

– *Информационные (основные) вопросы*, направленные на получение информации от каждого респондента.

– *Вопросы-фильтры*, используемые тогда, когда необходимы сведения не от всей совокупности респондентов, а только от некоторых. Это своеобразный опросник в опроснике. Начало и окончание фильтра обычно четко обозначают. Например: «Были ли за последний год госпитализации? Если да, пожалуйста, ответьте на следующие три вопроса, если нет, переходите к вопросу №...»

– *Контрольные вопросы* дают возможность уточнить правильность сообщенных респондентами сведений, а также исключить из дальнейшего рассмотрения недостоверные ответы или даже опросники. Среди контрольных вопросов можно выделить:

– повторения информационных вопросов, сформулированных другими словами. Если ответы основного и контрольного вопроса диаметрально противоположны, они исключаются из последующего анализа;

– вопросы, которые предлагают некоторое множество ответов там, где на практике может быть лишь однозначный ответ. Например: «Случалось ли вам капризничать в детстве?», «Всегда ли вы говорите правду?» и т. п., служат для выявления лиц, имеющих повышенную склонность к выбору социально одобряемых ответов.

Набор ответов (*response sets*) — тенденция респондентов отвечать на вопросы в форме, не соответствующей существу поставленных вопросов (Nunnally J. C., Bernstein I. H., 1994). Данный феномен является одной из проблем, связанных с использованием шкал. Чаще всего он проявляется в соглашательстве и социальной желательности.

Соглашательство (*поддакивание*) — тенденция с большей готовностью соглашаться, чем возражать. Особо актуальной данная проблема становится при обследовании лиц с интеллектуальными нарушениями.

Социальная желательность — предрасположенность или установка давать социально одобряемые ответы на самооценочные вопросы, чтобы создавать более благоприятное впечатление о себе самом у себя или окружающих. Искажающее влияние социальной желательности на полученные результаты особенно актуально в случае использования тестовых опросников, которые могут быть излишне «прозрачными». Для оценки выраженности феномена в конструкцию многих методик включены контрольные вопросы, совокупность которых формирует специальную оценочную шкалу, часто именуемую шкалой лжи. Например, в методике ММРІ шкала L включает те ут-

верждения, которые выявляют тенденцию испытуемого представить себя в возможно более выгодном свете, продемонстрировав очень строгое соблюдение социальных норм. Следует отметить, что опросники имеют разную «прозрачность». Кроме того, социальная желательность в той или иной степени свойственна не только опросникам, но и другим методам, в том числе многим проективным. Также неверно считать, что социальная желательность возникает и влияет на поведение только в ситуации тестирования. Скорее всего, данное качество обуславливает различные виды поведения индивида в различных ситуациях, и такого рода искажения могут сами по себе иметь некоторую диагностическую ценность. Снижению эффекта социальной желательности способствует правильное предъявление методики и предварительная беседа с испытуемым.

Правила, позволяющие повысить эффективность контроля:

— Основной и контрольный вопрос не следует размещать рядом, чтобы не была обнаружена их взаимосвязь.

— Ответы на прямые вопросы лучше контролировать косвенными вопросами.

— Контролю необходимо подвергать лишь наиболее существенные вопросы в анкете.

— Необходимость в контроле, как правило, снижается, если значительная часть вопросов допускает уклонения от ответа, выражения неопределенности мнения («не знаю», «затрудняюсь ответить», «когда как» и т. п.) (Голев С. В., Голева О. С., 2010а).

При подготовке опросника необходимо удостовериться в корректности формулировки вопросов. *В частности, исследователю нужно обратить внимание на следующие моменты:*

— Не содержит ли вопрос подсказки в явной или неявной форме? (Например, вопрос типа: «Что вам нравится в ...?» уже заранее предполагает, что нечто должно нравиться респонденту.)

— Не превышает ли вопрос уровня памяти или мышления респондента?

— Не присутствуют ли в вопросе слова, непонятные для респондента или имеющие крайне неопределенное содержание?

— Не задевает ли вопрос достоинства и самолюбия респондента?

— Не вызовет ли вопрос негативной эмоциональной реакции?

— Не слишком ли протяженный вопрос по размеру?

— Не являются ли ответы на вопрос излишне детализированными?

— Не спрашивается ли в вопросе одновременно о нескольких различных предметах?

— Нет ли ошибки в логике изложения?

— Относится ли вопрос ко всем потенциальным респондентам?

Необходимы ли фильтры?

- Нуждается ли вопрос в контроле? Если да, в каком именно?
 - Какой вид вопроса (по форме ответа и по формулировке) наиболее предпочтителен в данном конкретном случае?
 - Есть ли в закрытом вопросе варианты уклонения? Нужны ли они?
 - Имеется ли грамматическое согласование между вопросом и ответами к нему?
 - Не возникло ли искажений при перепечатке опросника?
- (Голев С. В., Голева О. С., 2010а).

3.2.2. Интервьюирование

Интервьюирование — метод получения информации в ходе устного непосредственного общения, который предусматривает регистрацию и анализ ответов на вопросы, а также изучение особенностей невербального поведения опрашиваемых. В отличие от обычной беседы, процедура интервьюирования имеет четкую цель, предполагает предварительное планирование действий по сбору информации, обработку полученных результатов (Голев С. В., Голева О. С., 2010б).

Достоинством интервьюирования является возможность его применения в широком диапазоне исследовательских целей. Это позволяет говорить об универсальности метода, а многообразие собираемых психологических фактов свидетельствует о значительном потенциале устного опроса.

Недостаток состоит в том, что интервьюирование считается одним из наиболее субъективных методов, поскольку в нем чрезвычайно велик риск получения недостоверного, преднамеренно или случайно искаженного сообщения. Прежде всего, искажения информации возникают из-за того, что респондент может отклониться от истины, если есть одна или несколько причин:

- податливость реальному или воображаемому давлению интервьюера;
- склонность к выражению социально одобряемых мнений;
- влияние на ответы имеющихся поведенческих установок и стереотипов;
- неотчетливое осознание собственных мнений, позиций и отношений;
- незнание каких-либо фактов или ложной информированности;
- антипатия к исследователю;
- сомнения в последующем сохранении конфиденциальности сообщения;
- сознательный обман или преднамеренное умолчание;
- невольные ошибки памяти (Голев С. В., Голева О. С., 2010б).

Источником искажений собираемых данных может стать и интервьюер.

Практика показывает, что устный опрос обеспечивает наилучшие результаты, если применяется в комплексе с другими средствами психологического исследования. В зависимости от условий проведения интервью данная процедура может быть единичной или многократной, индивидуальной или групповой. В зависимости от цели выделяют:

- собственно исследовательское интервью;
- диагностическое интервью, используемое на ранних этапах психотерапии как средство проникновения во внутренний мир пациента и понимания его проблем;
- клиническое интервью, являющееся, по сути, терапевтической беседой, способом оказания психологической помощи в осознании человеком внутренних затруднений, конфликтов, скрытых мотивов поведения, путей саморазвития личности.

По форме общения интервью подразделяются на свободные, структурированные (стандартизированные) и полуструктурированные (полустандартизированные) (Голев С. В., Голева О. С., 2010б; Newby P., 2014) (рис. 37).

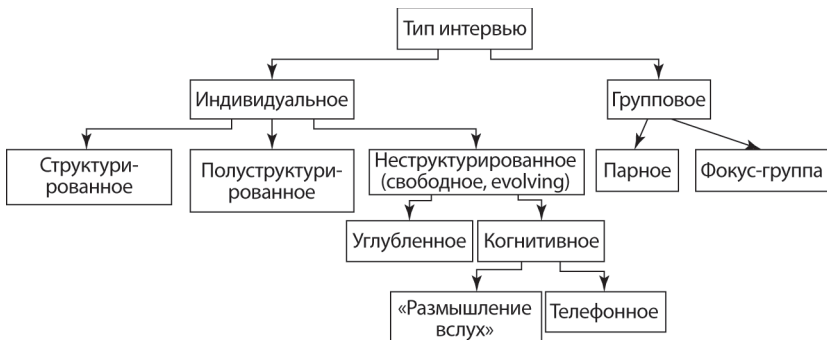


Рис. 37. Типы интервью (Newby P., 2014)

Свободное интервью представляет собой беседу, в которой исследователь имеет возможность самостоятельно изменять направленность, порядок и структуру вопросов, добиваясь необходимой эффективности процедуры. Характеризуется гибкостью построения диалога в пределах заданной темы, максимальным учетом индивидуальных особенностей респондентов, сравнительно большей естес-

твенностью условий опроса. *Достоинство* свободного интервью заключается в предоставлении респондентам наилучших возможностей для формулирования собственных точек зрения и более глубокого выражения позиции. Существенным *недостатком* является затрудненность сопоставлений всех полученных результатов, обусловленная широкой вариативностью задаваемых вопросов.

Структурированное интервью предусматривает проведение опроса по четко разработанной схеме, одинаковой для всех респондентов. Интервьюеру не разрешается изменять формулировки или порядок вопросов, задавать новые вопросы. Все условия процедуры регламентированы.

Достоинства:

- обеспечивается высокая степень сопоставимости всех индивидуальных результатов;

- сокращается до минимума количество ошибок при формулировании вопросов;

- повышается надежность итогов опроса.

Недостатки:

- мнения респондентов обычно не получают полного выражения;

- самому опросу присущ несколько формальный характер, затрудняющий достижение хорошего контакта между исследователем и опрашиваемыми лицами.

По сути, структурированное интервью мало отличается от опросников: разница состоит лишь в том, что заполнение опросников осуществляется самим исследователем в ходе беседы с респондентом.

Полуструктурированное интервью основано на использовании вопросов двух видов:

- обязательных, основных, которые должны быть заданы каждому опрашиваемому;

- уточняющих, которые применяются в беседе или исключаются из нее интервьюером в зависимости от ответов на основные вопросы.

Таким образом, достигается определенная вариативность опроса, учет индивидуальных особенностей респондентов и изменений коммуникативной ситуации. Полученные сведения сохраняют значительную сопоставимость. Исследователь активно управляет диалогом, при необходимости фокусируя внимание опрашиваемых на каких-либо дополнительных аспектах обсуждаемых проблем. Однако он не выходит за пределы заранее составленного перечня вопросов.

Независимо от разновидности интервьюирования существует определенная последовательность действий, обеспечивающая должную эффективность реализации данного метода (Голев С. В., Голева О. С., 2010б).

Подготовительный этап. Необходимо:

- определить предмет и объект опроса, исследовательские задачи, выбрать ту или иную разновидность интервью;

- отобрать инструментарий психологического исследования, в том числе составить план интервью, сформулировать примерный перечень вопросов респондентам, разработать инструкции, подготовить технические средства регистрации и обработки данных;

- провести пилотажное интервьюирование;

- уточнить программу исследования, отредактировать вопросы, внести в случае необходимости изменения в инструкции, проанализировать ошибки и несоответствия, возникшие в ходе пробного интервью;

- составить отредактированную версию перечня вопросов, текста инструкции для респондентов.

Требования к вопросам интервью во многом совпадают с требованиями, предъявляемыми к вопросам опросников.

Коммуникативный этап. Проводится непосредственная беседа интервьюера и респондента.

Структура общения с респондентом:

- введение в беседу: установление контакта, информирование о целях опроса и условиях его проведения, формирование установки на сотрудничество, ответы на возникающие у респондента вопросы;

- основная фаза интервью: подробное исследование, осуществляемое по заранее разработанному плану;

- завершение беседы: снятие возникшего напряжения, выражение благодарности и признательности за участие в работе (Голев С. В., Голева О. С., 2010б).

Успех интервью во многом зависит от того, насколько доброжелательным и заинтересованным собеседником окажется исследователь. Вступительное слово должно быть кратким, обоснованным и уверенным. Сообщение о цели исследования излагается в той форме, которая стимулирует респондента на совместную работу. Манера общения интервьюера должна быть в целом нейтральной. Иногда интервьюер сталкивается с нежеланием респондента отвечать на тот или иной вопрос, в этом случае исследователь должен отнестись к нему с уважением, даже если лишается важной информации. Однако представляется вполне возможным возвращение к уже затронутой ранее теме на более поздних этапах опроса и в иной формулировке. В целях сохранения естественности условий опроса интервьюер стремится к тому, чтобы задавать большинство вопросов по памяти, не прибегая к своим записям. Не стоит тратить время на изучение плана или припоминание очередной темы. Наличие такого рода затруднений нередко стимулирует респондента на попытки перехвата инициативы,

превращения интервью в обычную беседу (Голев С. В., Голева О. С., 2010б).

Фиксация информации в интервью представляет особую проблему. Возможно использование технических средств регистрации ответов (диктофон, видеокамера), что позволит в случае необходимости повторно вернуться в записи беседы и проанализировать ее нюансы. Восстановление текста беседы на основе медиафайлов возможно при помощи специализированных компьютерных программ, например: EXMARaLDA Partitur-Editor (www.emeraldaorg.org) (Schmidt T., 2011). Однако маскировка таких технических средств не соответствует этическим принципам исследования. Открытая же запись с помощью видеокамеры, диктофона или магнитофона, как и стенографирование хода интервью или ведение исследователем дословных записей, приводит к тому, что респонденты чувствуют себя весьма стесненно, дают искаженные ответы. В то же время фиксация информации только по памяти, по окончании процедуры опроса, зачастую приводит к целому ряду существенных искажений материала.

Аналитический этап. Производятся обработка и интерпретация собранной информации, ее анализ, результаты устного опроса сопоставляются с данными, полученными при помощи других методов психологического исследования.

Специфическим подходом к анализу информации, полученной в ходе интервью, является углубленное исследование текста, зафиксированного непосредственно в ходе беседы или восстановленного (транскрибированного) по записи беседы техническими средствами. Для анализа этого материала используются методы, получившие обобщающее название **дискурсивный анализ** (*discourse analysis*) (Barker С. [et al.], 2002; Newby P., 2014), одним из вариантов которого является контент-анализ.

Дискурсивный анализ предполагает изучение текста с особым вниманием к его структуре, содержанию, определенным словам и формулировкам, их сочетаниям, понятиям, речевым оборотам, метафорам. Например, задача может состоять в оценке того, какие слова используют пациенты, когда они говорят о психических заболеваниях. Текстовые материалы могут быть сформированы не только в ходе интервью, но и в процессе изучения естественных взаимодействий (например, диалога детей). Объектом анализа могут быть тексты, специально сформированные по просьбе и при помощи исследователя. Например, исследователь может попросить пациента написать о тех трудных жизненных ситуациях, с которыми ему пришлось столкнуться, описать переживания, связанные с заболеванием, лечением и т. п.

Для реализации дискурсивного анализа предложен ряд прикладных компьютерных программ. К их числу можно отнести программы GATE (<https://gate.ac.uk/>) (Cunningham H. [et al.], 2015), CATMA (www.slm.uni-hamburg.de/catma) (Schüch L., 2010), AQUAD (<http://www.aquad.de/en/eng/index.html>) (Huber G. L., Gürtler L., 2015) и ряд других.

Пример

Анализ имитации беседы врача и пациента (Соложенкин В. В., 2011) проведен при помощи программы CATMA. Выявлено, что в своем рассказе пациент чаще всего употребляет слова, связанные с понятием «боль», при этом статистически достоверно чаще эти слова употребляются в комбинации со словом «голова», что укладывается в описание симптомов болезни пациентом, страдающим артериальной гипертензией (рис. 38, 39, см. цв. вкл.).

3.2.3. Наблюдение

Наблюдение не только возможно в любых исследованиях и обстоятельствах, но и неизбежно. Даже если в качестве основного способа изучения объекта применяется какой-либо другой эмпирический метод, наблюдение обязательно сопровождает его и является неотъемлемой частью обследования (Голев С. В., Голева О. С., 2010а). Так, при тестировании и опросе, психофизиологических измерениях и психотерапевтических воздействиях, сборе биографической информации и социально-психологических исследованиях исследователь наблюдает за реакциями и поведением испытуемого, следит за соблюдением правил и условий проведения обследования.

Наблюдение может выступать и в качестве специального методического приема со своей спецификой и особенностями. В подобной ситуации **наблюдение** — целенаправленное, организованное и фиксируемое восприятие психических явлений с целью изучить их в определенных условиях. Следовательно, к научному психологическому наблюдению предъявляются определенные требования: наличие цели, организованность процесса наблюдения и регистрация получаемых данных, полнота и адекватность фиксируемого материала изучаемым явлениям.

Преимущество наблюдения состоит в том, что это непосредственный способ фиксирования поведения испытуемого, что позволяет получить конкретную информацию об интересующем исследователя феномене.

Пример

Целью исследования является изучение взаимодействия семейной пары. При анкетировании супругов можно получить ответы, свидетельствующие о том, что они не могут нормально общаться и что их беседы обычно заканчиваются ничем. Однако при непосредственном наблюдении за ними можно выявить конкретную проблему, приводящую к нарушению их коммуникации.

Пример

Исследуются нарушения поведения ребенка. Родители могут жаловаться на то, что ребенок непослушен. При непосредственном наблюдении можно увидеть, что же на самом деле происходит при взаимодействии родителей и ребенка.

Кроме того, наблюдение позволяет оценивать поведение в определенном контексте. Так, наблюдение за ребенком в школе и дома позволяет исследователю выявить и оценить возможные ситуативные переменные (например, критические высказывания брата или сестры в адрес ребенка, критические высказывания одноклассников), которые могут повлиять на поведение. Наблюдение также полезно при изучении аспектов поведения, которые сам испытуемый не осознает, а также тех особенностей поведения, которые он некритично оценивает вследствие отрицания, искажения информации или простого забывания.

Недостатки:

— Наблюдение может позволить дать ответы только на определенные вопросы исследования, особенно те, которые связаны с непосредственной фиксацией ряда феноменов. Однако часто задачи исследования более сложные, и внешние поведенческие проявления — лишь один из аспектов проблемы.

— Сказывается реактивность измерения: люди могут вести себя совершенно иначе, если знают, что за ними наблюдают. Выходом из этой ситуации может быть скрытое наблюдение, что, очевидно, связано с этическими проблемами.

Методологически наблюдение может осуществляться с использованием как качественных, так и количественных оценок. В случае *качественных оценок* исследователь составляет своеобразный отчет, который, подобно литературному произведению, позволяет воссоздать картину того, что было увидено. В отличие от литературного произведения этот отчет должен быть предельно эксплицитным и систематическим, а не метафорическим и интуитивным.

Основные ситуации, когда применяется целенаправленное научное наблюдение:

— ориентировка в проблеме (получение материала, способствующего прояснению проблемы, уточнению вопросов, формулировке гипотез);

— основной сбор данных об объекте и предмете исследования, когда другие методы неприемлемы или их применение затруднено;

— дополнение, уточнение или контроль результатов, полученных другими методами;

— иллюстрация предложенных теорий, гипотез, догадок (Голев С. В., Голева О. С., 2010а).

Наблюдение как метод имеет ряд модификаций (Голев С. В., Голева О. С., 2010а).

Объективное (внешнее) наблюдение — наблюдение внешних по отношению к наблюдателю объектов. Как самостоятельный способ исследований применяется во всех разделах психологии, в том числе и клинической.

Самонаблюдение (интроспекция) — наблюдение субъекта за самим собой, за актами собственного сознания и поведения. В качестве главного элемента самонаблюдение входит в состав аутогенной тренировки, психологического тренинга, где, по инструкции, требуется следить за своими переживаниями и поведением. В опосредованном виде самонаблюдение проявляется в документальных источниках, подлежащих психологическому анализу: в письмах, дневниках, автобиографиях.

Полевое (естественное) наблюдение — наблюдение за объектами в естественных условиях их повседневной жизни и деятельности. В полной мере может быть реализовано в форме объективного наблюдения. Естественное наблюдение предстает обычно как самостоятельный способ сбора данных. Его совмещение с другими эмпирическими методами возможно, когда наблюдение выступает ведущим, основным приемом, а другие методы — вспомогательными, сопутствующими и разнесенными с ним во времени. Одновременное их применение неизбежно влияет на естественный ход жизнедеятельности изучаемых объектов, и тогда говорить об этом виде наблюдения уже не приходится.

Лабораторное (экспериментальное) наблюдение осуществляется в искусственно создаваемых условиях. Степень этой искусственности может быть различной. Минимум искусственности достигается в непринужденной беседе и в привычной для испытуемого обстановке, а максимум — в условиях эксперимента с использованием технических средств и инструкций в специальных помещениях. Этот вид наблюдения почти всегда связан с применением других эмпирических методов. В медицинской практике данная разновидность наблюдения часто именуется **клиническим наблюдением**, т. е. наблюдением за пациентом в процессе его лечения.

Индивидуальное наблюдение осуществляется одним наблюдателем. Этот наблюдатель может выступать как единственным исследователем в рамках конкретного исследования, так и одним из группы исследователей. В последнем случае он все равно выполняет функции единственного исследователя, но уже в рамках какого-либо этапа или раздела общего исследования.

Коллективное наблюдение осуществляется совместно несколькими наблюдателями. При этом совместность определяется в первую

очередь общностью исследования (единый план, цель, методика), а не единством места и времени работы наблюдателей, хотя обычно предполагается, что за одним и тем же объектом одновременно наблюдают несколько исследователей.

Случайное наблюдение — не запланированное заранее наблюдение, совершаемое в силу неожиданно сложившихся обстоятельств. Наблюдение ценно прежде всего в отношении редких явлений, предугадать которые невозможно. Важно, чтобы исследователь был заранее готов к таким ситуациям.

Преднамеренное наблюдение — заранее запланированное, входящее в замысел исследователя и преследующее определенные цели. Именно преднамеренные, а не случайные наблюдения составляют основную массу научных сведений.

Систематическое наблюдение — преднамеренное наблюдение, совершаемое по заранее обдуманному плану и, как правило, по заранее составленному графику. Систематичность выражается в процедурном и временном аспектах. *Процедурный аспект* подразумевает ясное представление целей и задач наблюдения, четкое формулирование рабочей гипотезы, определенность и упорядоченность действий наблюдателя, продуманность всей системы регистрируемых показателей поведения наблюдаемых объектов и условий среды. Последний фактор иногда считается основополагающим для данного вида наблюдения, особенно если регистрируемые показатели выражены в определенных единицах и категориях описания. *Временной аспект* заключается в планировании и сбалансированности многократных наблюдений, преследующих одну и ту же цель. При этом многократные наблюдения могут быть направлены на один и тот же или на разные объекты, осуществляться одним или несколькими исследователями, входить в один или несколько исследовательских циклов.

Несистематическое наблюдение — нестрогое наблюдение без определенного плана. *Процедурный аспект* сказывается в неопределенности задач наблюдения, отсутствии заранее предусмотренных признаков, подлежащих регистрации; в отсутствии четкого алгоритма действий наблюдателей. *Временной аспект* выражается в случайности совершения многократных наблюдений, хотя сами единичные (одноразовые) наблюдения могут быть и систематическими.

Полное наблюдение охватывает и фиксирует максимум доступной наблюдателю информации. Применяется для предельно тщательного изучения объекта. Нередко полное наблюдение осуществляется как вынужденная мера в случаях, когда заранее неизвестно, какие факторы ситуации и поведения следует регистрировать.

Неполное наблюдение — наблюдение, при котором внимание исследователя обращается на оптимальное число параметров. Круг подлежащих регистрации сведений определяется заранее, исходя из задач и условий наблюдения.

Сплошное наблюдение представляет собой постоянное наблюдение за объектом без перерыва. Обычно применяется при краткосрочном изучении или сборе наиболее полной информации о динамике изучаемых явлений.

Выборочное наблюдение проводится в отдельные промежутки времени, выбираемые исследователем по своему усмотрению. Оно экономичнее сплошного и целесообразно при длительных исследованиях.

Констатирующее наблюдение состоит в том, что замечаемые явления и действия лишь фиксируются и не подлежат обсуждению или оцениванию исследователем в ходе наблюдения. Как правило, эти зарегистрированные факты имеют однозначное толкование, не требующее текущей интерпретации.

Разновидностью констатирующего наблюдения является фотографическое, сопровождающееся подробнейшей регистрацией всего наблюдаемого без высказывания наблюдателем своего отношения к происходящему.

Оценивающее наблюдение сопровождается вынесением наблюдателем оценки ситуации или фиксируемых явлений и фактов.

Стандартизованное наблюдение осуществляется по заранее разработанной схеме, предписывающей форму фиксации и перечень подлежащих регистрации параметров. Часто для этого используются бланки наблюдения. Наблюдение применяется, когда изучаемый процесс или явление в целом понятны и требуется лишь отследить их элементы, уточнить детали, получить дополнительный материал. Результат — четкость и сравнимость данных, возможность количественных оценок.

Нестандартизованное наблюдение — нерегламентированное наблюдение, при котором описание происходящего производится наблюдателем в свободной форме.

Наблюдение может быть **открытым**, при котором наблюдаемые знают о своей роли объекта исследования, обычно им известен наблюдатель. Относительная свобода действий исследователя сочетается с некоторыми трудностями психологического характера. Раскрыв цели исследования, наблюдатель рискует лишиться сотрудничества или повлиять на дальнейшее поведение людей, которое может стать неестественным. Эффекты особенно усиливаются, если испытуемые знают о наблюдении, но им неизвестен наблюдатель. Это вызывает у них беспокойство, дискомфорт, попытки обнаружить наблюда-

теля, а следовательно, искажает обычное поведение наблюдаемого (Голев С. В., Голева О. С., 2010а).

При **скрытом наблюдении** испытуемым о факте наблюдения не сообщается, и наблюдение проводится незаметно для них. Несмотря на определенные этические проблемы, связанные со скрытым наблюдением, главное его преимущество состоит в отсутствии влияния наблюдателя на испытуемых.

Включенное наблюдение (*participant observation*). Исследователь внедряется в организацию или социальную группу для получения информации из первых рук, изучения изнутри. Предусматривается период интенсивного социального взаимодействия между исследователем и теми людьми, за которыми он наблюдает, в естественных для испытуемых условиях, при этом данные собираются систематично и беспрепятственно, хотя и в неструктурированной форме (Taylor S. J., Bogdan R., 1998). Авторы обращают внимание на опасность проведения включенного наблюдения в тех случаях, когда исследователь лично знаком с участниками исследования и тем учреждением, в котором проводится наблюдение. Это связано с неизбежным соблазном цензурировать те находки, которое могут как-то дискредитировать коллег или друзей. Включенное наблюдение вызывает ряд вопросов о надежности и валидности такого метода. Действительно, весьма затруднительно дать оценку точности тех оценок, которые дал исследователь. Теоретически можно одновременно задействовать сразу нескольких наблюдателей и сопоставить их данные, но практически это осуществляется редко. Оценки, даваемые исследователем, могут быть сильно искажены, так как на них могут оказать влияние его предшествующие надежды и ожидания. Это искажение влияет на то, как исследователь видит происходящее, воспроизводит увиденное. Кроме того, присутствие наблюдателя среди обследуемых лиц создает риск реактивности испытуемых: факт наблюдения влияет на поведение тех людей, за которыми осуществляется наблюдение. Такая проблема не является уникальной для исследований, основанных на качественных оценках, она характерна для всех видов наблюдений. Исследователь может снизить этот эффект, выделив достаточное время для привыкания испытуемых к его присутствию. Некоторые исследователи пытаются избежать этой проблемы, организовав скрытое наблюдение, что, в свою очередь, сопряжено с этическими вопросами.

Невключенное наблюдение осуществляется со стороны, без взаимодействия наблюдателя с объектом изучения. Этот вид наблюдения, по сути, есть объективное (внешнее) наблюдение.

Прямое (непосредственное) наблюдение проводится непосредственно самим наблюдателем.

Косвенное (опосредованное) наблюдение проводится через посредников. Обычно имеет в виду получение данных от свидетелей и участников событий. Опосредованное наблюдение может предполагать использование технических средств (например, видео- или звукозаписывающей аппаратуры).

Спровоцированное наблюдение состоит в том, что исследователь провоцирует наблюдаемого на какие-либо действия. Этот прием употребляется обычно для акцентирования изучаемого явления и для его совпадения во времени с процедурой наблюдения.

Неспровоцированное наблюдение предполагает отсутствие специального провоцирующего влияния на ход событий.

В целях стандартизации процедуры возможно привнесение в наблюдение элементов *количественных оценок*, например подсчета числа эпизодов, сопровождавшихся теми или иными поведенческими актами, продолжительности того или иного типа поведения. При этом необходимо четкое операционное определение поведения того типа, которое необходимо зафиксировать и учесть, хотя это не всегда легкая задача. Оцениваемые показатели и методы их фиксации должны быть заранее четко и недвусмысленно определены. В ходе работы подготовленный исследователь использует четко сформулированные «коды», соответствующие определенным вариантам поведения. В качестве примера приводится кодировка вариантов поведения, использовавшаяся исследователями при наблюдении за взаимодействием учителя и учеников (табл. 5). Как правило, количественный подход к оценке результатов наблюдения ориентирован на оценки ограниченных поведенческих актов. Следует также заранее определить, какие аспекты будут учитываться в количественной форме (например, типы актов агрессии, их частота, интенсивность).

Пример

Наблюдение направлено на обнаружение проявлений агрессии у детей во время игр. Заранее четко определено, что рассматривать как акт агрессии, а что агрессией не является, т. е. задан «код» агрессии (так, если один ребенок дотронулся до другого во время игры, то при каких обстоятельствах можно считать, что он его толкнул, и при каких — ударил?).

Фрагмент кодировки отдельных аспектов поведения учителя
(Bolhuis S., Voeten M., 2001; Barker C. [et al.], 2002)

Категория наблюдения	Описание категории
1. Объяснение материала	Учитель говорит, и предполагается, что ученики его слушают
e1 Обсуждение темы	Учитель объясняет, уточняет, демонстрирует и освещает вопросы, связанные с изучаемой на уроке темой
e2 Привлечение внимания	Учитель привлекает внимание учеников к изучаемой на уроке теме
e3 Повторение пройденного материала	Учитель напоминает ученикам материал, разбиравшийся на предыдущих уроках
e4 Введение	Учитель рассказывает ученикам о том, что будет изучаться на этом и/или следующих уроках
e5 Релевантность	Учитель объясняет ученикам, какое значение имеет изучаемая тема для практической, внешкольной жизни
e6 Оценка	Учитель объясняет, как будут оцениваться результаты изучения учениками темы
e7 Обучение	Учитель рассказывает, как выучить тему, как выполнить задания и проверить правильность их выполнения
e8 Цели обучения	Учитель объясняет и демонстрирует ученикам цели обучения
2. Вопросы	Учитель задает ученикам вопросы
q1 Относящиеся к теме занятия	Учитель задает вопросы по теме занятия
q2 Повторение пройденного материала	Учитель задает вопросы по пройденному материалу
q3 Проверка	Учитель задает вопросы, чтобы убедиться в том, что материал понят и можно продолжить разбор темы («Все ли вам понятно?», «Все поняли?»)
q4 Учебная работа	Учитель спрашивает, как ученики выполняли задания и как планируют это делать
q5 Диагностика/мониторинг	Учитель просит учеников подумать, проанализировать и оценить свою учебу

Категория наблюдения	Описание категории
q6 Релевантность	Учитель спрашивает учеников о значении темы
q7 Повтор	Учитель повторяет вопрос, если на него не был получен правильный ответ
q8 Продолжение	Учитель дополняет вопрос, чтобы уточнить знания ученика
q9 Общий вопрос	Учитель задает вопрос, непосредственно не относящийся к теме занятия и процессу обучения

3.3. Оценка качества измерений

3.3.1. Надежность измерения

Надежность — воспроизводимость тех измерений, которые были получены при помощи данной методики (Barker С. [et al.], 2002). Чем более стабильны выявляемые результаты, тем больше доверия вызывает методика и тем меньше величина ошибки, которая влияет на процесс измерения. Математически это можно определить как долю вариации измеряемого параметра, связанной с ошибкой измерения, в общей вариации измеряемого признака. Высокая надежность методики позволяет измерить показатели с достаточной уверенностью и выявить связь между изучаемыми признаками, которая может быть не распознана, если надежность мала. Если методика абсолютно ненадежна, то она регистрирует случайную ошибку, а не истинные показатели. Уровень надежности еще не говорит о том, что же реально измеряет данная методика: высокая надежность лишь свидетельствует о повторяемости получаемых результатов. Например, термометр, из которого вытекла жидкость, будет всякий раз показывать одинаковые цифры, т. е. давать надежные измерения, хотя они и будут неверными. Значение, верность измеренных показателей определяется уже валидностью измерений.

Выбор конкретного подхода к оценке надежности зависит от метода исследования (наблюдение, анкетирование), шкалы, используемой при измерениях, а также ряда других обстоятельств. Однако чаще всего для оценки надежности используются следующие приемы:

- внутреннее согласие (в том числе принцип разделения пополам);
- межисследовательское согласие;

- повторное тестирование;
- эквивалентные формы опросников (Barker С. [et al.], 2002).

Оценка внутреннего согласия (*internal consistency*) – стандартный подход к оценке взаимной согласованности признаков, составляющих шкалу опросника. Предполагается, что признаки, входящие в опросник, эквивалентны или параллельны, т. е. нацелены на изучение одного и того же конструкта. Внутреннее согласие позволяет оценить, измеряют ли признаки одно и то же (высокое внутреннее согласие) или разные вещи (низкое внутреннее согласие). Чаще всего изучается взаимная ковариация признаков, входящих в опросник, что статистически находит свое выражение в показателе α **Кронбаха** (*Cronbach's alpha*). Расчет проводится следующим образом:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S_T^2} \right),$$

где k – число признаков, входящих в опросник; S_i – среднеквадратическое отклонение показателей i -го признака ($i = 1, \dots, k$), S_T – среднеквадратическое отклонение суммарных показателей по всем признакам опросника.

Если признаки, входящие в опросник, абсолютно не связаны по смыслу друг с другом, α равно нулю; если же все признаки измеряют фактически одно и то же, α будет равно 1,0. Принято считать, что внутреннее согласие удовлетворительное при величине $\alpha > 0,7$. Показатели менее 0,5 считаются неудовлетворительными и требуют пересмотра методики (Woodward М., 2014). Очевидно, не только низкий, но и чрезмерно высокий ($> 0,9$) показатель α не может расцениваться как удовлетворительная находка, поскольку косвенно свидетельствует о тавтологии вопросов, входящих в опросник.

Пример

Рассчитывается показатель α Кронбаха на примере собственных данных оценки постнатальной депрессии при помощи Эдинбургской шкалы (Edinburgh Postnatal Depression Scale, EPDS). Опросник состоит из 10 вопросов, представленных в форме шкалы Лайкерта. Сумма баллов по всем вопросам отражает выраженность постанатальной депрессии. В табл. 6 представлены данные обследования 82 женщин. Расчет выглядит следующим образом:

$$k = 10; \sum S_i^2 = 6,078; S_T^2 = 21,842;$$

$$\alpha = [10/(10 - 1)] \cdot (1 - 6,078/21,842) = 0,802.$$

Полученный показатель свидетельствует о высоком уровне внутреннего согласия.

Расчет показателя α Кронбаха можно выполнить при помощи большинства прикладных статистических программ (SPSS, STATA, SAS и др.). В ходе таких расчетов возможно осуществление ряда дополнительных важных процедур, например поочередного удаления из опросника отдельных признаков с оценкой соответствующего изменения показателя α Кронбаха: если при удалении определенного вопроса показатель Кронбаха увеличивается и внутренняя согласованность улучшается, данный вопрос, скорее всего, является лишним для опросника и может быть исключен.

Таблица 6

**Данные обследования 82 женщин по Эдинбургской шкале
постнатальной депрессии, балл**

Наблю- дение	Признак по Эдинбургской шкале										Общий показа- тель
	№ 1	№ 2	№ 3	№ 4	№ 5	№ 6	№ 7	№ 8	№ 9	№ 10	
1	2	1	2	3	2	2	0	0	0	0	12
2	3	3	3	3	3	3	0	3	3	2	26
3	2	0	2	2	2	2	0	0	0	0	10
4	1	1	3	3	3	2	2	2	2	0	19
5	0	0	1	1	0	0	0	0	0	0	2
6	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	1	1	0	0	1	0	0	3
8	1	0	1	1	0	1	0	0	0	0	4
9	0	0	1	2	1	0	0	0	0	0	4
10	0	0	2	2	2	0	1	1	1	0	9
11	1	0	0	0	0	3	0	0	1	0	5
12	0	0	2	0	2	2	0	0	0	0	6
13	0	1	1	0	1	1	0	0	0	0	4
14	1	0	0	0	0	0	1	0	1	0	3
15	0	0	0	1	0	0	0	0	0	0	0
16	0	1	2	2	1	1	0	1	1	0	9
17	0	0	1	2	1	1	0	1	1	0	7
18	1	0	2	2	1	2	0	1	1	0	10
19	1	0	2	1	1	2	0	0	0	0	6
20	0	0	0	1	0	1	0	0	0	0	2
21	1	2	1	1	0	1	0	1	0	0	7
22	2	0	0	1	2	1	0	1	2	0	9
23	0	0	2	0	0	1	0	0	0	0	3
24	0	0	2	0	0	0	0	0	0	0	2

Наблю- дение	Признак по Эдинбургской шкале										Общий показа- тель
	№ 1	№ 2	№ 3	№ 4	№ 5	№ 6	№ 7	№ 8	№ 9	№ 10	
25	2	1	2	2	2	2	2	1	1	0	15
26	0	0	0	1	1	0	0	1	0	0	3
27	1	0	1	1	0	1	0	0	0	0	4
28	1	0	0	0	3	0	0	0	1	0	5
29	0	0	0	0	0	0	0	0	0	0	0
30	1	1	3	3	2	2	1	1	0	0	14
31	0	2	0	2	2	1	0	0	1	0	8
32	0	0	1	1	0	1	2	0	0	0	5
33	0	0	2	0	2	1	0	0	0	0	5
34	0	0	0	2	1	1	0	1	0	0	5
35	0	0	0	0	0	0	0	0	1	0	1
36	1	0	0	0	0	1	0	1	0	0	3
37	0	1	1	1	2	2	1	1	0	0	9
38	0	1	1	2	1	1	1	1	1	0	9
39	1	0	1	1	1	0	0	1	1	0	6
40	0	0	0	1	1	1	0	0	0	0	3
41	0	0	2	2	2	0	0	0	1	0	7
42	0	0	1	1	0	0	0	0	0	0	2
43	1	1	1	2	0	1	0	1	0	0	7
44	0	0	3	2	2	2	0	1	1	1	12
45	0	0	1	1	0	1	0	0	1	0	4
46	0	0	1	0	0	1	0	0	0	0	2
47	0	0	2	1	0	0	0	0	0	0	3
48	0	1	3	2	2	2	0	1	1	0	12
49	0	0	1	0	0	0	0	0	1	0	3
50	2	0	2	1	0	2	2	2	1	0	12
51	1	1	1	2	0	0	1	1	1	0	8
52	0	0	2	2	2	1	0	0	1	0	8
53	1	1	2	1	2	1	0	0	1	0	10
54	0	0	1	3	3	2	0	1	1	0	10
55	0	0	2	2	2	0	0	1	0	0	7
56	0	0	0	2	0	6	0	0	1	0	5
57	1	0	1	1	1	1	0	0	0	0	5
58	0	1	0	0	0	0	0	0	0	0	1

Наблю- дение	Признак по Эдинбургской шкале										Общий показа- тель
	№ 1	№ 2	№ 3	№ 4	№ 5	№ 6	№ 7	№ 8	№ 9	№ 10	
59	0	2	2	1	1	2	1	2	1	0	12
60	0	0	1	2	2	1	0	1	1	0	8
61	0	0	1	0	0	0	0	0	0	0	1
62	0	0	0	1	0	0	0	1	1	0	3
63	1	0	0	0	0	0	0	0	0	0	1
64	0	1	2	0	1	0	0	0	0	0	4
65	1	1	0	0	0	1	0	0	0	0	3
66	0	1	2	2	1	1	0	0	1	0	8
67	0	0	0	0	0	0	0	0	0	0	0
68	0	1	1	1	2	1	3	1	1	0	11
69	0	0	0	1	0	0	0	0	0	0	1
70	0	0	1	0	0	1	0	0	0	0	2
71	2	0	0	2	0	1	0	0	0	0	5
72	1	2	2	2	1	2	1	0	1	0	12
73	0	0	2	0	0	0	0	1	0	0	3
74	1	0	2	3	0	3	2	0	1	0	12
75	0	1	0	0	0	0	0	0	0	0	1
76	0	1	0	1	0	1	0	0	0	0	3
77	1	0	2	0	0	1	0	1	0	0	5
78	1	1	1	1	1	1	1	1	0	0	8
79	0	0	2	0	1	0	0	0	0	0	3
80	0	2	2	0	1	1	3	3	3	0	14
81	0	0	2	0	2	1	0	0	0	0	5
82	0	0	0	0	0	0	0	0	0	0	0
S ²	0,473	0,441	0,883	0,908	0,892	1,010	0,486	0,475	0,450	0,060	21,842 (S _T ²)
	ΣS _i ² = 6,078										

Расчеты изменений внутреннего согласия при удалении отдельных вопросов, выполненные при помощи программы SPSS, представлены в табл. 7. Ни один исключенный признак не улучшает показателя внутреннего согласия. Между отдельно взятыми признаками и оставшимися показателями наблюдается хорошая корреляция, удаление признаков существенно не снижает общей вариации значений шкалы. Таким образом, можно сделать вывод о том, что все признаки, включенные в шкалу, хорошо взаимно согласованы.

**Влияние отдельных признаков шкалы на характеристики
внутреннего согласия**

Признак	Среднее значение шкалы при удалении признака	Вариация показателя шкалы при удалении признака	Скорректированная корреляция признака с оставшимися показателями шкалы	Значение α Кронбаха при удалении признака
Без удаления признаков	6,17	22,119	–	0,802
1	5,72	19,315	0,386	0,799
2	5,77	18,896	0,482	0,789
3	5,04	17,246	0,511	0,786
4	5,10	16,731	0,575	0,777
5	5,32	17,083	0,531	0,783
6	5,22	17,136	0,477	0,792
7	5,87	19,303	0,380	0,799
8	5,70	18,116	0,602	0,777
9	5,68	18,145	0,616	0,776
10	6,13	21,056	0,445	0,803

Метод разделения пополам в настоящее время используется реже. Вопросы, входящие в опросник, делятся на две эквивалентные формы (например, содержащие четные и нечетные вопросы). Затем определяется корреляция показателей, полученных по двум формам опросника. Такой анализ проведен применительно к рассматриваемой Эдинбургской шкале: первую половину составили вопросы 1–5, вторую – вопросы 6–10. Показатель α Кронбаха составил 0,697 для первой половины и 0,668 для второй половины. Корреляция между показателями, полученными по двум половинам шкалы, – 0,658. Поскольку коэффициент корреляции рассчитывается между двумя половинами шкалы, т. е. при числе признаков меньшем, чем в исходном опроснике, в вычисленный показатель следует вносить поправку, так как при прочих равных условиях при большем количестве признаков, входящих в опросник, показатель внутреннего согласия возрастает. Используют формулу предсказания Спирмена–Брауна (*Spearman–Brown prophecy formula*):

$$r_{kk} = \frac{k \cdot r_{11}}{1 + (k - 1) \cdot r_{11}},$$

где r_{kk} – показатель надежности опросника при объединении всех признаков; k – коэффициент, отражающий кратность увеличения количества признаков (в нашем случае он равен 2, так как суммарное число признаков увеличится в 2 раза); r_{11} – исходный показатель (Barker C. [et al.], 2002; Reddy M. V., 2015).

Таким образом, формула приобретает вид:

$$r_{22} = 2r_{11}/(1 + r_{11}).$$

Скорректированный показатель r_{22} равен 0,794 и говорит о высокой внутренней согласованности признаков, входящих в опросник, и надежности методики.

Воспроизведение результатов исследования, полученных различными исследователями у одних и тех же испытуемых, потенциально может подвергнуться влиянию разных факторов:

– *Биологическая вариабельность* изучаемых показателей у одного и того же обследуемого; примером могут служить показатели артериального давления, исследуемые у одного и того же пациента при различных обстоятельствах, в различное время суток и т. п.

– Методика исследования изначально может иметь высокую вероятность вариабельности показателей, полученных в ходе повторных измерений. Чаще всего это обусловлено *техническими погрешностями методики*.

— Факторы, прямо или косвенно связанные с людьми, выполняющими диагностические исследования. При этом различают **внутриисследовательскую вариабельность** (*intra-observer variability*), сводящуюся к тому, что один и тот же исследователь при повторном изучении показателей у одного и того же испытуемого может выявлять различные результаты, а также **межисследовательскую вариабельность** (*inter-observer variability*), которая заключается в неодинаковых подходах к использованию методики и трактовке ее результатов различными исследователями. Межисследовательская вариабельность сводится к минимуму в тех случаях, когда четко регламентирована процедура проведения исследования и заданы критерии, на основании которых возможна интерпретация результатов. Эти условия легче обеспечить, если используются количественные оценки. Задача осложняется, если используются качественные оценки, особенно при высокой вероятности субъективизма трактовок результатов.

Уровень межисследовательского согласия (*inter-observer consistency*) выражается в том, в какой мере различные исследователи однозначно трактуют результаты применения методики (Кельмансон И. А., 2002; Власов В. В., 2004; Barker С. [et al.], 2002). Статистические подходы к оценке межисследовательского согласия во многом зависят от используемой шкалы измерений. Поскольку в исследованиях довольно часто используются номинальные шкалы, нужно подробно рассмотреть статистические оценки, применяемых в этих случаях. Предположим, что методика предполагает вынесение бинарного суждения, например, о наличии или отсутствии заболевания, патологического состояния и т. п. Пример распределения оценок при вынесении таких суждений двумя независимыми исследователями представлен в табл. 8. Полное совпадение оценок исследователей наблюдается в $a + b$ случаях, и степень межисследовательского согласия p_0 рассчитывается следующим образом:

$$p_0 = (a + d) / (a + b + c + d).$$

Нередко значение степени согласия приводят в процентах. Однако такая оценка учитывает совпадения, связанные как с одинаковыми подходами исследователей к трактовке результатов, так и с чистой случайностью.

Таким образом, истинное суждение о межисследовательском согласии должно предполагать коррекцию на возможность случайного совпадения результатов. С указанной целью используется **критерий каппа (κ) Коэна** (*Cohen's kappa*) (Armitage P. [et al.], 2001; Barker С. [et al.], 2002). Его расчет сводится к следующему. Прежде всего определяют фактическую (observed) степень согласия p_0 по вышеп-

риведенной формуле. Затем рассчитывают ожидаемую (calculated) степень согласия p_c , исходя из гипотезы о том, что такое совпадение определяется исключительно влиянием случая. Расчет p_c выглядит следующим образом:

$$p_c = \frac{(a + c) \cdot (a + b) + (b + d) \cdot (c + d)}{n^2}.$$

Критерий κ рассчитывается следующим образом:

$$\kappa = \frac{p_0 - p_c}{1 - p_c}.$$

Таблица 8

Распределение заключений (бинарные оценки) по данным двух независимых исследователей

Исследователь 1	Исследователь 2		Суммарно
	Положительные	Отрицательные	
Положительные	a	b	$a + b$
Отрицательные	c	d	$c + d$
Суммарно	$a + c$	$b + d$	$n = a + b + c + d$

Значение критерия κ может достигать 1,0 при полном совпадении результатов, а его минимально возможная величина равна нулю при полном несовпадении. Предложены различные подходы к интерпретации значений критерия κ (Szklo M., Nieto F. J., 2007) (рис. 40). Чаще всего об отличном межисследовательском согласии говорят при $\kappa \geq 0,75$, об удовлетворительном при $0,4 < \kappa < 0,75$; о плохом при $\kappa \leq 0,4$ (Fleiss J. L. [et al.], 2003).

При большом числе обследованных ($n > 100$) можно достаточно точно оценить среднеквадратическое отклонение, соответствующее критерию κ (Власов В. В., 2004):

$$\sigma_k = \sqrt{\frac{p_0(1 - p_0)}{n(1 - p_c)^2}}.$$

Тогда статистическая достоверность критерия κ будет оцениваться на основе z -преобразования:

$$z = \kappa / \sigma_k.$$

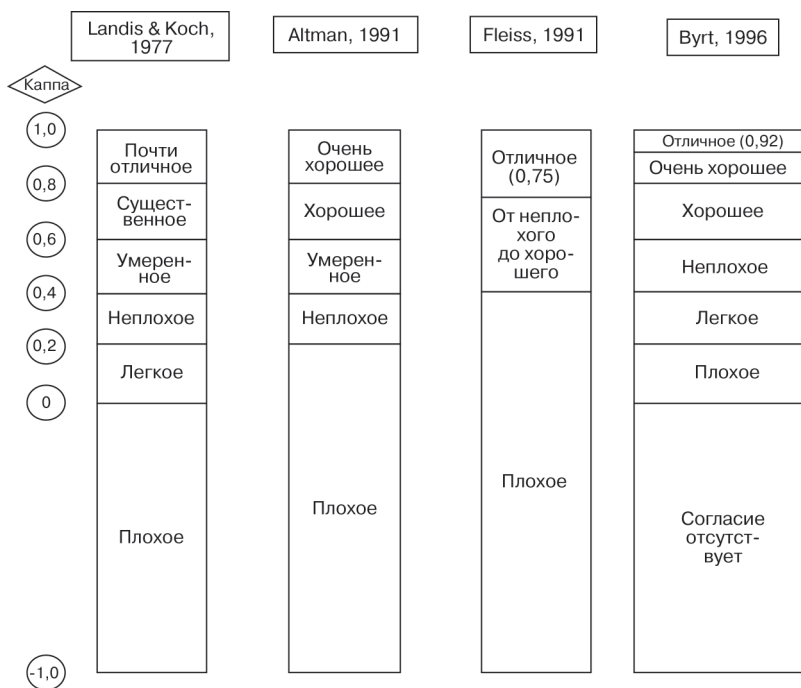


Рис. 40. Различные подходы к интерпретации значений коэффициента Козна (Szklo M., Nieto F. J., 2007)

Для оценки уровня статистической значимости полученное значение z необходимо сопоставить с таблицей граничных значений для t -критерия.

Когда проводится сопоставление результатов теста, предполагающего возможность более двух оценок, расчеты несколько усложняются. Если число возможных оценок равно l , то мера согласия будет основываться на разнице фактических и теоретически ожидаемых значений наблюдений в диагональных ячейках табл. 9. Ожидаемое число наблюдений в каждой ячейке таблицы будет рассчитываться следующим образом:

$$E_{ii} = \frac{R_i \cdot C_i}{n},$$

где E_{ii} — ожидаемое число наблюдений в ячейке (i, i) в строке i и столбце i .

**Представление данных для оценки степени согласия оценок
двух исследователей при 4 возможных категориях
показателя (А – D)**

Исследователь 1	Исследователь 2				Суммарно
	А	В	С	Д	
А	O11	O12	O13	O14	R1
В	O21	O22	O23	O24	R2
С	O31	O32	O33	O34	R3
Д	O41	O42	O43	O44	R4
Суммарно	C1	C2	C3	C4	n

Таким образом, согласие может быть рассчитано как сумма разностей фактического числа и теоретически ожидаемого числа наблюдений в диагональных ячейках таблицы:

$$\sum_{i=1}^l (O_{ii} - E_{ii})$$

Пример

Два исследователя должны выявить у 100 испытуемых наличие диагноза шизофрении, аффективных нарушений или иных психических расстройств (Barker С. [et al.], 2002) (табл. 10). Фактическое совпадение диагнозов, установленных двумя исследователями, составит: $10 + 20 + 20 = 50$. Таким образом, степень согласия p_0 составляет $50/100 = 0,5$.

Таблица 10

**Гипотетическое распределение диагнозов у 100 обследованных
пациентов на основании оценок двух независимых исследователей
(Barker С. [et al.], 2002)**

Исследователь 1	Исследователь 2			Суммарно
	Шизофрения	Аффективные расстройства	Прочие нарушения	
Шизофрения	10	20	0	30
Аффективные расстройства	10	20	10	40
Прочие нарушения	0	10	20	30
Суммарно	20	50	30	100

Для вычисления p_c необходимо прежде всего рассчитать долю наблюдений, приходящихся на каждую строку и каждый столбец таблицы: это дела-

ется суммированием числа наблюдений в соответствующих строках и столбцах и делением полученных значений на общее число наблюдений. Затем вычисляется p_c путем последовательного перемножения значений долей, приходящихся на соответствующие строку и столбец для каждой клетки таблицы, и суммирования полученных значений:

$$p_c = 0,3 \times 0,2 + 0,4 \times 0,5 + 0,3 \times 0,3 = 0,06 + 0,20 + 0,09 = 0,35.$$

Тогда значение k составит:

$$k = (0,5 - 0,35)/(1 - 0,35) = 0,23.$$

В приведенном примере, основанном на использовании номинальной шкалы признаков, возможно изучение показателя надежности и для какой-то отдельно взятой категории этой шкалы. Для этого категории признака нужно свести к двум: той, которая представляет интерес для анализа, и всем остальным суммарно. Так, может стоять вопрос: насколько надежной является диагностика шизофрении? Тогда таблица будет сведена к двум категориям: «диагноз шизофрении» и «отсутствие диагноза шизофрении», т. е. будет сведена к виду 2×2 , что позволит выполнить соответствующие расчеты показателя k .

Если для измерения признака используются не номинальные, как в приведенном примере, а ординальные и интервальные шкалы, оценка показателя надежности усложняется.

Ординальная шкала с числом градаций признака более двух. В принципе возможен расчет показателя k в соответствии с приведенной выше методикой. Однако в случае ординальной шкалы расхождение оценок, данных двумя исследователями, будет гораздо более существенным, если они дают сильно различающиеся, а не близкие, хотя и различные оценки признака, и было бы разумным принять во внимание это обстоятельство при исследовании межисследовательского согласия. Например, при оценке выраженности депрессивной симптоматики возможна не просто бинарная оценка «да/нет», а оценка на ординальной шкале: «нет симптомов депрессии — умеренная выраженность симптомов — сильная выраженность симптомов». Очевидно, если один исследователь диагностирует у испытуемого отсутствие симптомов депрессии, а второй — их умеренную выраженность, то такое расхождение оценок не столь критично; более значимым будет расхождение, если второй исследователь диагностирует у этого испытуемого сильную выраженность симптомов депрессии. Учесть высказанные соображения позволяет расчет взвешенного показателя k Козна (Armitage P. [et al.], 2001; Woodward M., 2014). Анализируются не только диагональные, но все ячейки таблицы с учетом весового значения показателей. Тогда степень межисследовательского согласия p_0 составит:

$$p_0 = \frac{\left(\sum_{i=1}^l \sum_{j=1}^l w_{ij} O_{ij}\right)}{n},$$

а ожидаемое согласие p_c :

$$p_c = \frac{\left(\sum_{i=1}^l \sum_{j=1}^l w_{ij} R_i \cdot C_j\right)}{n^2},$$

При этом вес каждой ячейки (i, j) определяется следующим образом:

$$w_{ij} = 1 - \frac{|i - j|}{l - 1},$$

где $|i - j|$ — абсолютное значение (модуль) разности $(i - j)$ (знак разности при этом не учитывается).

Пример

Изучаются данные о различных оценках социальной принадлежности обследованных женщин (Woodward M., 2014). При этом выделено 6 социальных классов (число категорий $l = 6$) (табл. 11). Представлены и соответствующие каждой ячейке весовые значения (табл. 12).

Значение p_0 составит:

$$p_0 = (1 \times 22 + 0,8 \times 0 + 0,6 \times 73 + 0,4 \times 6 + \dots + 1 \times 97) / 3334 = 0,83.$$

Значение p_c составит:

$$p_c = (1 \times 194 \times 36 + 0,8 \times 194 \times 841 + \dots + 1 \times 185 \times 344) / 3334^2 = 0,70.$$

Таким образом, взвешенное значение κ составит:

$$\kappa = (0,83 - 0,70) / (1 - 0,70) = 0,43.$$

Таблица 11

Распределение обследованных женщин по социальным классам в результате двух оценок (Woodward M., 2014)

Оценка 1	Оценка 2						Суммарно
	I	II	III n	III m	IV	V	
I	22	80	73	6	9	4	194
II	11	471	241	31	60	12	826
III n	3	61	379	20	29	15	507
III m	0	159	326	197	263	152	1097
IV	0	60	92	43	266	64	525
V	0	10	26	11	41	97	185
Суммарно	36	841	1137	308	668	344	3334

I, II, III n, III m, IV, V – различные социальные классы.

Веса w_{ij} для ячеек, представленных в табл. 11

Оценка 1	Оценка 2					
	I	II	III n	III m	IV	V
I	1,0	0,8	0,6	0,4	0,2	0
II	0,8	1,0	0,8	0,6	0,4	0,2
III n	0,6	0,8	1,0	0,8	0,6	0,4
III m	0,4	0,6	0,8	1,0	0,8	0,6
IV	0,2	0,4	0,6	0,8	1,0	0,8
V	0	0,2	0,4	0,6	0,8	1,0

Еще более сложной является задача исследования межисследовательского согласия в случае интервальной шкалы измерений. Если на интервальной шкале выбрать какой-то пороговый показатель, позволяющий отнести обследуемого к определенной категории (например, пороговый уровень по шкале депрессии Бека для констатации депрессивной симптоматики), то можно трансформировать полученные результаты в бинарную номинальную форму: «есть депрессия» — «нет депрессии». Тогда для оценки межисследовательского согласия можно осуществлять расчет показателя k . Если интервальные шкалы имеют непрерывные значения показателя, вопрос о межисследовательском согласии решается с использованием коэффициента корреляции Пирсона r между показателями, зафиксированными первым и вторым исследователями.

Однако исследование парных корреляций для оценки степени межисследовательского согласия таит в себе определенную опасность. Действительно, если один исследователь при оценке признака дает стойко завышенные или заниженные показатели (например, на 10 баллов выше, чем первый исследователь), то коэффициент парной корреляции показателей первого и второго исследователя будет равен 1,0. Однако такая «идеальная» корреляция вряд ли будет свидетельствовать о хорошем межисследовательском согласии.

Вот почему более содержательным подходом к оценке межисследовательского согласия является исследование **внутриклассовой корреляции** (*intraclass correlation, ICC*), которая позволяет учесть различия в полученных оценках, связанные с исследователями. Наиболее распространенным подходом является допущение, согласно которому исследователи, дающие оценки одним и тем же субъектам, отбираются случайно. Модель описывается следующим уравнением (Howell D. C., 2013):

$$X_{ij} = \mu + \alpha_i + \pi_j + \alpha\pi_{ij} + e_{ij},$$

где X_{ij} — значение изучаемого признака, полученное i -м исследователем у j -го испытуемого; α_i — эффект, связанный с i -м исследователем; π_j — эффект, связанный с j -м испытуемым; $\alpha\pi_{ij}$ — эффект взаимодействия i -го исследователя с j -м испытуемым (эффект, отражающий то, как мнение конкретного исследователя меняется при обследовании конкретного испытуемого); e_{ij} — ошибка, возникающая при обследовании i -м исследователем j -го испытуемого.

Оценка выраженности эффекта измерения обычно предполагает исследование показателя вариации признака, связанной с различными источниками происхождения. Именно на этом принципе, базирующемся на дисперсионном анализе (ANOVA), и основан расчет коэффициента внутриклассовой корреляции:

$$ICC = \sigma^2\pi / (\sigma^2\alpha + \sigma^2\pi + \sigma^2\alpha\pi + \sigma_e^2).$$

Если основная доля вариации значений показателя связана с вариацией между отдельными испытуемыми $\sigma^2\pi$, а вариации, связанные с различиями между отдельными исследователями $\sigma^2\alpha$, с взаимодействием «исследователь — испытуемый» $\sigma^2\alpha\pi$, с ошибкой измерения σ_e^2 невелики, тогда значение коэффициента внутриклассовой корреляции будет близко к 1,0. Интерпретация значений коэффициента внутриклассовой корреляции в целом аналогична интерпретации значений коэффициента корреляции.

Предложен и более простой способ расчета коэффициента внутриклассовой корреляции:

$$ICC = \frac{S_A^2 + S_B^2 - S_{diff}^2}{S_A^2 + S_B^2 + \bar{X}_{diff}^2 - \frac{S_{diff}^2}{n}},$$

где S_A — это среднеквадратическое отклонение значений показателя, измеренных исследователем А; S_B — среднеквадратическое отклонение значений показателя, измеренного исследователем В; S_{diff} — среднеквадратическое отклонение разницы значений показателей, измеренных исследователями А и В; \bar{X}_{diff} — среднее значение разницы показателей, полученных исследователями А и В; n — число наблюдений (Szklo M., Nieto F. J., 2007).

Пример

Рассмотрим данные об исследовании числа реакций пробуждений за 1 ч времени сна (индекс пробуждений) двумя исследователями (А и В), выполненном у 30 пациентов (Szklo M., Nieto F. J., 2007) (табл. 13). Расчет коэффициента внутриклассовой корреляции выглядит следующим образом:

$$ICC = (56,7 + 123,02 - 73,31)/(56,7 + 123,02 + (-3,67)^2 - 73,31/30) = 0,56.$$

Таблица 13

**Сопоставление результатов исследования индекса:
число пробуждений за 1 ч сна, осуществленного двумя исследователями
(Szklo M., Nieto F. J., 2007)**

Номер наблюдения	Исследователь А	Исследователь В	Разница (А – В)
1	7,08	8,56	-1,48
2	18,60	19,91	-1,31
3	20,39	25,17	-4,78
4	16,39	22,68	-6,29
5	27,95	17,21	10,74
6	29,57	27,15	2,42
7	13,50	18,66	-5,16
8	24,50	20,58	3,92
9	14,63	15,61	-0,98
10	11,15	13,10	-1,95
11	19,52	19,05	0,47
12	18,91	18,59	0,32
13	17,98	10,78	7,20
14	15,78	12,64	3,14
15	0,00	7,04	-7,04
16	8,15	10,75	-2,60
17	20,36	20,61	-0,25
18	36,62	34,90	1,72
19	18,31	20,84	-2,53
20	17,56	24,28	-6,72
21	8,14	22,94	-14,80
22	17,30	19,38	-2,08
23	16,39	22,68	-6,29
24	29,29	65,09	-35,80
25	18,80	18,75	0,05
26	10,92	20,97	-10,05
27	12,53	13,38	-0,85

Номер наблюдения	Исследователь А	Исследователь В	Разница (А – В)
28	24,94	43,92	-18,98
29	18,66	18,02	0,64
30	12,50	23,25	-10,75
Среднее	17,55	21,22	-3,67
S	7,53	11,09	8,56
S ²	56,70	123,02	73,31

Методика Бланда–Альтмана (*Bland–Altman plot*) (Altman D. G., 1999; Woodward M., 2014) является весьма интересным подходом к оценке согласованности двух измерений количественных показателей. Она представляет собой графическое отображение зависимости разности двух измеренных показателей от среднего значения этих показателей. Такой анализ позволяет определить, имеется ли систематическая ошибка измерения, вследствие которой различия между измерениями зависят от уровня изучаемого показателя. Если точки на графике равномерно распределяются выше и ниже горизонтальной линии, проходящей через нулевую отметку (т. е. линию, соответствующую отсутствию разницы показателей), а границы ДИ вокруг среднего значения достаточно узкие, систематические различия в оценках двух исследователей отсутствуют. Если же точки преимущественно расположены выше или ниже нулевой линии или имеется тенденция (тренд) к перемещению точек вверх или вниз по мере увеличения среднего показателя, можно говорить о наличии систематической ошибки двух оценок. Выявление достоверности тренда сдвига точек по мере роста среднего значения показателей можно оценить при помощи коэффициента корреляции тау (τ) Кендалла (Peat J. K., Barton B., 2014). На рис. 41 представлен график Бланда–Альтмана, основанный на данных измерения индекса пробуждений двумя исследователями (см. табл. 13). В целом исследователь А давал более низкие оценки показателя (в среднем на 3,7); отмечается довольно широкий разброс разницы значений: с 95 %-й вероятностью разница показателей, полученных двумя исследователями, находилась в интервале от 13,1 до -20,5. Однако не выявлен тренд в разнице показателей в зависимости от среднего значения (коэффициент корреляции Кендалла $\tau = 0,03$; $P = 0,82$).

Методика Бланда–Альтмана может быть использована и для сопоставления результатов измерений одного и того же показателя, полученных на основе различных методов. В собственном исследовании было проведено изучение качества жизни, связанного со здоровьем,

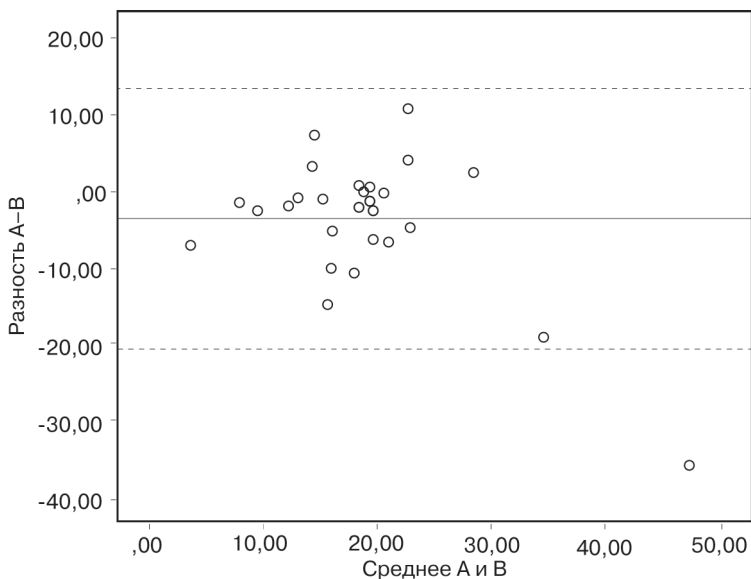


Рис. 41. Сопоставление результатов исследования индекса: число пробуждений за 1 ч сна, осуществленного двумя исследователями: А и В (Szklo M., Nieto F. J., 2007). Сплошная горизонтальная линия — среднее значение разности показателей, пунктирные горизонтальные линии — границы 95 %-го ДИ разности

у детей старшего дошкольного возраста, страдающих бронхиальной астмой. Оценки проводились на основании опросника KINDL, причем детям было предложено заполнить детскую форму опросника, а родителям — взрослую. Обе формы дают оценку качества жизни по 100-балльной шкале. Цель — изучить согласие в оценке качества жизни детей, даваемой самими детьми и их родителями. На рис. 42 представлен график Бланда—Альтмана, отражающий полученные результаты. В целом дети были склонны давать более низкие оценки качества жизни (средняя разность $-2,04$ балла). Имелся весьма широкий диапазон различий показателей, полученных у детей и взрослых (границы 95 %-го ДИ составили от $23,68$ до $-27,76$ баллов), что свидетельствовало о неудовлетворительной согласованности оценок детей и родителей. Анализ графика также выявляет тренд: смещение точек в область более высоких значений по мере роста среднего показателя качества жизни, полученного на основе оценок детей и родителей. Этот тренд был статистически достоверным (τ Кендалла = $0,23$; $P = 0,045$). По мере роста усредненного показателя качества

жизни дети склонны давать более высокие оценки по сравнению с родителями, а по мере снижения среднего показателя качества жизни родители склонны выше оценивать качество жизни ребенка, чем это делает сам ребенок.

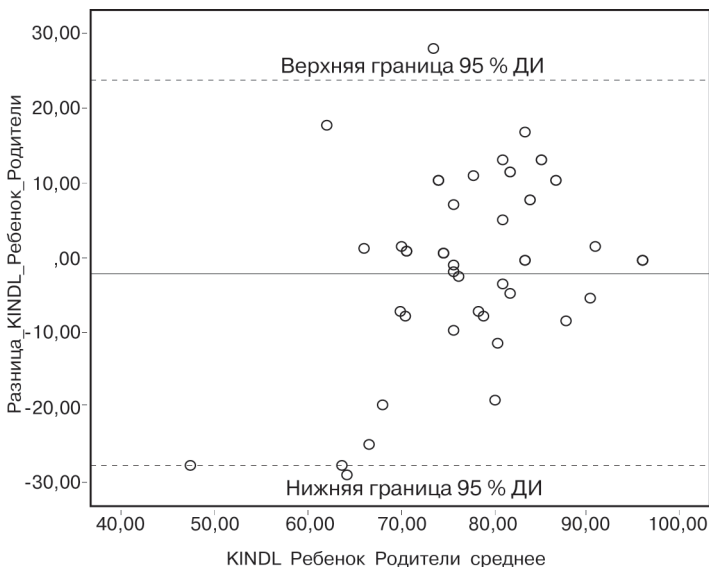


Рис. 42. Оценка качества жизни (опросник KINDL) у детей старшего дошкольного возраста с диагнозом «бронхиальная астма»

В тех ситуациях, когда проводится изучение межисследовательского согласия при участии более чем двух исследователей, в качестве меры такого согласия можно использовать показатель α Кронбаха, подобно тому как это осуществляется при исследовании внутреннего согласия с той лишь разницей, что в данном случае каждый отдельный исследователь выступает в роли одного из признаков опросника. Также мерой межисследовательского согласия может выступать и коэффициент внутриклассовой корреляции.

Повторное тестирование (*test-retest*). Соответствующая оценка надежности методики отражает устойчивость показателей в течение определенного времени. Измерение проводится в одной и той же выборке, но несколько раз (2 или более). Временной промежуток между двумя измерениями может составлять несколько дней, неделю, месяц. Надежность в ходе повторного тестирования часто обозначается как

коэффициент стабильности. При использовании повторного тестирования на результаты могут повлиять различные факторы, возникающие как во время самих опросов (например: различные условия заполнения опросников при первом и повторных опросах; влияние самого опроса; специфические факторы, связанные с состоянием респондента; продолжительность периода между опросами и т. д.), так и между ними. Например, значения могут существенно измениться, если больной первый раз заполнил опросник при выраженной симптоматике и при повторном визите, когда на фоне назначенного лечения симптоматика была купирована. В данном случае при оценке воспроизводимости будет получен низкий показатель: поскольку не учтены факторы, влияющие на состояние респондента, это не означает, что опросник недостаточно надежен (Новик А. А., Ионова Т. И., 2007).

Статистические подходы, используемые для оценки коэффициента стабильности, будут принципиально те же, что при изучении межисследовательского согласия, с той лишь разницей, что будут сопоставляться показатели, полученные одним и тем же исследователем, но в ходе повторных испытаний.

Эквивалентные формы (*equivalent forms*). Принцип позволяет оценивать надежность методики путем сопоставления результатов, полученных при использовании различных версий одного и того же диагностического инструмента. Модификация принципа повторного тестирования выражена в том, что при повторном тестировании вместо использования той же методики применяют ее альтернативную («эквивалентную», «параллельную») версию. В некоторых психодиагностических методиках заранее предусмотрены формы «А» и «В», что упрощает процедуру применения эквивалентных форм.

3.2.2. Валидность измерения

Валидность измерения (*measurement validity*) является ответом на вопрос, «измеряет ли измерение то, что оно должно измерить» (Barker С. [et al.], 2002). Например, измеряет ли шкала для оценки симптомов депрессии выраженность депрессии или же она измеряет что-то другое (самооценку, желание признать наличие проблемной ситуации и т. п.)? Сам по себе концепт «валидность измерения» более сложен для понимания и оценки, чем надежность. В настоящем разделе обсуждается именно валидность измерения, хотя постановка вопроса о валидности распространяется и на дизайн исследования, о чем говорилось в главе 2. Изучение качества методики измерения того или иного показателя предполагает, что на первом этапе требуется оценка надежности, а затем валидности. С одной стороны, на-

дежность является необходимым, но не достаточным условием валидности. Для того чтобы быть валидным, измерение должно быть надежным, в противном случае оно будет преимущественно отражать влияние ошибки. Так, если два исследователя не могут прийти к единому суждению по поводу того, содержатся ли в тексте беседы пациента и психотерапевта признаки отказа пациента от лечения, нет возможности делать какие-то выводы о валидности категории отказа от лечения. С другой стороны, методика может быть очень надежной, но не валидной: например, измерение окружности головы для определения интеллекта. Оцениваются различные аспекты валидности методики, на которые нужно ориентироваться исследователю. К основным видам валидности следует отнести содержательную, внешнюю, критериальную и конструктивную.

Содержательная валидность (*content validity*) — показатель того, насколько методика (а также каждый ее компонент или шкала) охватывает все аспекты того конструкта, который методика призвана измерить. Например, присутствуют ли в опроснике с целью выявить симптомы депрессии у пациента вопросы, затрагивающие сниженное настроение, снижение мотивации, нарушения сна и т. д. Оценка содержательной валидности носит качественный характер, и нет какого-то специального коэффициента содержательной валидности. Имеется три способа оценки содержательной валидности:

- анализ общепринятых подходов к изучению рассматриваемой проблемы на основании литературных данных;
- экспертная оценка содержания опросника специалистами;
- оценка содержания опросника самими больными.

Внешняя валидность (*face validity*) весьма близка к содержательной валидности. Оценка внешней валидности предполагает ответ на вопрос: лежит ли на поверхности то, что призвана измерить данная методика? Так, в опроснике, цель которого выявить симптомы депрессии, должны присутствовать вопросы, затрагивающие снижение настроения у пациента, а не его отношение к начальнику на работе. Внешняя валидность позволяет определить, охватывают ли вопросы опросника круг проблем, беспокоящих пациента. Внешняя валидность желательна, но не всегда обязательна. Так, в опроснике могут присутствовать «мелкие» на первый взгляд, не относящиеся к изучаемому конструкту вопросы, которые на самом деле позволяют снизить вероятность фальсификации данных. Внешняя валидность отчасти является своеобразным пиар-ходом, так как в результате шкала опросника выглядит правильной для потенциального респондента. Если у респондента не складывается такого впечатления, он может просто отказаться отвечать на поставленные вопросы. Как и содержательная валидность, внешняя валидность является качественной оценкой и не имеет какого-то измерения.

Критериальная валидность (*criterion validity*) устанавливает возможность точно оценить ту или иную переменную или тот или иной концептуальный показатель на основе сопоставления с «золотым стандартом» и возможность использования методики для замены такого «золотого стандарта». Оценка критериальной валидности методики предполагает использование и сравнение эффективности двух сопоставляемых методик (одновременное или попеременное в случайной последовательности), причем важно, чтобы разрабатываемая методика была проще в использовании и экономичнее, чем существующий метод оценки, рассматриваемый в качестве стандарта (Кельмансон И. А., 2004). Проблема критериальной валидности во многом является центральной в решении вопроса о валидности методики. Критериальная валидность позволяет оценить, насколько удовлетворительно измеренный показатель согласуется с другим показателем, принятым в качестве критерия, индикатора измеряемого конструкта. Такой принятый критерий часто именуется «золотым стандартом» диагностики. Существуют разновидности критериальной валидности:

— *Конкурентная валидность*: диагностика осуществляется на основе принятого критерия одновременно с анализируемой методикой. Результаты использования анализируемой методики сопоставляются с существующими критериями. Например, оценки по шкале симптомов депрессии могут быть сопоставлены с результатами обследования пациента врачом-психиатром.

— *Предсказательная валидность*: диагностика осуществляется в дальнейшем, когда оценивается реализация прогноза, предсказания, основанного на анализируемом методе. Оценка по шкале депрессии может быть рассмотрена в качестве прогностического фактора последующих суицидальных попыток.

В отличие от других видов валидности, критериальная имеет ряд количественных характеристик. Критериальная валидность диагностической методики прежде всего определяется тем, насколько эффективно она способна констатировать позитивные результаты теста у тех обследуемых, которые в действительности имеют распознаваемое заболевание или высокий риск его развития, и негативные — у лиц, не имеющих данного заболевания или характеризуемых низким риском его развития. В табл. 14 представлено возможное соотношение результатов диагностического теста и истинного наличия или отсутствия распознаваемого заболевания (состояния), верифицированного «золотым стандартом» диагностики. В этой таблице величине a соответствует число обследуемых, имеющих положительные результаты диагностического теста при истинном наличии заболевания (**истинно положительные результаты**), величине b — число обследуе-

мых, имеющих положительные результаты теста при фактическом отсутствии заболевания (**ложноположительные результаты**), величине c — число обследуемых, имеющих отрицательные результаты диагностического теста при истинном наличии заболевания (**ложноотрицательные результаты**), величине d — число обследуемых, имеющих отрицательные результаты диагностического теста при отсутствии заболевания (**истинно отрицательные результаты**).

Таблица 14

Сопоставление данных диагностического теста и «золотого стандарта» диагностики

Результат диагностического теста	Наличие заболевания (по данным «золотого стандарта»)		Всего
	Да	Нет	
Позитивные	a	b	$a + b$
Негативные	c	d	$c + d$
Суммарно	$a + c$	$b + d$	$a + b + c + d$

Двумя принципиальными операционными характеристиками теста, отражающими его критериальную валидность, являются чувствительность (*sensitivity*, Se) и специфичность (*specificity*, Sp) (Кельмансон И. А., 2002; Власов В. В., 2004; Кельмансон И. А., 2004; Покровский В. П., Брико Н. И., 2012; Barker С. [et al.], 2002; Rothman К. J. [et al.], 2008).

Чувствительность — вероятность того, что результаты диагностического теста окажутся положительными при истинном наличии заболевания. Показатель вычисляется следующим образом:

$$Se = a / (a + c).$$

По мере увеличения чувствительности теста снижается число испытуемых, фактически имеющих заболевание и неверно классифицируемых как здоровые в результате диагностического теста (снижается число ложноотрицательных результатов). Иначе говоря, если тест имеет высокую чувствительность, его негативные результаты позволяют эффективно исключить наличие диагноза.

Для правильной интерпретации смысла показателя чувствительности предлагается мнемоническое правило: **SnNout** (если тест имеет высокую чувствительность (**Sensitivity**), его отрицательные (**Negative**) результаты позволяют отвергнуть (**rule out**) диагноз) (Кельмансон И. А., 2002, 2004; Straus S. E., 2011).

Специфичность — вероятность того, что результаты теста окажутся отрицательными при истинном отсутствии заболевания у обследуемого. Специфичность рассчитывается следующим образом:

$$Sp = d / (b + d).$$

Высокая специфичность теста означает низкую вероятность его положительных результатов у лиц с отсутствием заболевания (низкую вероятность ложноположительных результатов). Иначе говоря, если тест имеет высокую специфичность, его позитивный результат позволяет эффективно распознать диагностируемое заболевание. Правильно интерпретировать смысл показателя специфичности помогает мнемоническое правило: **SpPin** (если тест характеризуется высокой специфичностью (**Sp**ecificity), его позитивные (**Pos**itive) результаты подтверждают (rule **in**) диагноз) (Кельмансон И. А., 2002, 2004; Straus S. E., 2011).

Пример

Рассмотрим данные валидизации Эдинбургской шкалы постнатальной депрессии (Choi S. K. [et al.], 2012). Было обследовано 250 женщин; в качестве порогового уровня диагностики высокого уровня депрессивной симптоматики авторы выбрали значение ≥ 3 баллов по шкале (табл. 15).

Таблица 15

Результаты исследования валидности Эдинбургской шкалы постнатальной депрессии (Choi S. K. [et al.], 2012)

Результат диагностического теста	Наличие симптомов депрессии		Всего
	Да	Нет	
Позитивные	$a = 35$	$b = 26$	$a + b = 61$
Негативные	$c = 3$	$d = 186$	$c + d = 189$
Суммарно:	$a + c = 38$	$b + d = 212$	$a + b + c + d = 250$

Расчет показателей чувствительности и специфичности дает следующие результаты:

$$Se = a / (a + c) = 35 / 38 = 0,92 \text{ (92 \%)},$$

$$Sp = d / (b + d) = 186 / 212 = 0,88 \text{ (88 \%)}.$$

При выбранном пороговом уровне тест имеет высокую чувствительность, равную 92 %. Следовательно, лишь в 8 % случаев существует вероятность того, что женщины, фактически имеющие проявления депрессивной симптоматики, будут неверно классифицироваться как здоровые. Тест имеет и достаточно высокую специфичность: в 88 % случаев положительные результаты теста позволяют правильно диагностировать наличие депрессивной симптоматики.

Очевидно, было бы весьма желательно, чтобы тест одновременно характеризовался высокой чувствительностью и специфичностью, как в приведенном примере. Однако на практике такая ситуация встречается редко, и можно говорить о существовании конкурентных отношений: по мере увеличения чувствительности наблюдается снижение специфичности теста и наоборот. Такая взаимоисключающая связь показателей чувствительности и специфичности отчасти связана с тем, что диагностика заболевания предполагает наличие лиц с доказанным отсутствием этого заболевания, доказанным наличием, а также существование некой «серой зоны», в пределах которой констатация наличия или отсутствия заболевания зависит от строгости используемых критериев. В подобных случаях снижение порогового уровня (выбран для констатации заболевания), направленное на снижение риска просмотра заболевания (а значит, на повышение чувствительности теста), должно неизбежно приводить к увеличению доли тех лиц, у которых будет ошибочно диагностироваться заболевание, т. е. к снижению специфичности.

Пример

Предположим, имеется гипотетическая методика, направленная на выявление определенного диагноза. Получены результаты обследования 30 испытуемых по данной методике, выявлено наличие у них распознаваемого диагноза (табл. 16). Очевидно, на операционные характеристики данного диагностического теста будет влиять выбор того порогового уровня, после достижения которого можно констатировать наличие заболевания. Зависимость чувствительности и специфичности теста от порогового уровня суммы баллов представлена в табл. 17. Снижение порогового уровня диагностики приводит к возрастанию чувствительности теста, но создает высокую вероятность ложноположительных результатов за счет неизбежного снижения специфичности. Напротив, завышение порога суммы баллов, необходимой для констатации диагноза, приводит к возрастанию вероятности пропуска тех лиц, которые в действительности имеют диагноз; в то же время доля ложноположительных результатов снижается.

Таблица 16

Распределение баллов по гипотетической диагностической шкале и наличия диагноза у 30 испытуемых

Порядковый номер испытуемого	Балл по диагностической шкале	Наличие диагноза
1	10	Нет
2	15	Нет
3	12	Нет
4	30	Нет

Порядковый номер испытуемого	Балл по диагностической шкале	Наличие диагноза
5	11	Нет
6	14	Нет
7	20	Нет
8	25	Нет
9	9	Нет
10	13	Нет
11	28	Нет
12	40	Нет
13	9	Нет
14	8	Нет
15	14	Нет
16	35	Да
17	42	Да
18	48	Да
19	50	Да
20	11	Да
21	15	Да
22	50	Да
23	45	Да
24	32	Да
25	48	Да
26	52	Да
27	60	Да
28	14	Да
29	20	Да
30	36	Да

Зависимость чувствительности и специфичности диагностики от порогового значения, выбранного для установления диагноза

Пороговое значение для установления диагноза (равно или больше приведенного значения)	Чувствительность Se	Специфичность Sp	$0,5Se + 0,5Sp$	Индекс Юдена ($Se + Sp - 1$)
7,0	1,00	0,00	0,50	0,00
8,5	1,00	0,07	0,53	0,07
9,5	1,00	0,20	0,60	0,20
10,5	1,00	0,27	0,63	0,27
11,5	0,93	0,33	0,63	0,27
12,5	0,93	0,40	0,67	0,33
13,5	0,93	0,47	0,70	0,40
14,5	0,87	0,60	0,73	0,47
17,5	0,80	0,67	0,73	0,47
22,5	0,73	0,73	0,73	0,47
26,5	0,73	0,80	0,77	0,53
29,0	0,73	0,87	0,80	0,60
31,0	0,73	0,93	0,83	0,67
33,5	0,67	0,93	0,80	0,60
35,5	0,60	0,93	0,77	0,53
38,0	0,53	0,93	0,73	0,47
41,0	0,53	1,00	0,77	0,53
43,5	0,47	1,00	0,73	0,47
46,5	0,40	1,00	0,70	0,40
49,0	0,27	1,00	0,63	0,27
51,0	0,13	1,00	0,57	0,13
56,0	0,07	1,00	0,53	0,07
61,0	0,00	1,00	0,50	0,00

Решение о выборе порогового уровня показателей теста, необходимого для принятия решения, во многом зависит от конкретной ситуации, прежде всего от значимости риска недооценки, пропуска распознаваемого состояния. Так, вполне обоснованным можно считать повышение чувствительности теста за счет снижения его специфичности, если речь идет о диагностике тяжелого заболевания, которое нельзя пропустить, а также в тех случаях, когда проведение теста не требует больших материальных затрат и технически легко реали-

зуюмо (например, измерение артериального давления). Увеличение специфичности обосновано в случаях дорогостоящих и труднореализуемых тестов. Таким образом, чувствительности и специфичности может придаваться разный вес в зависимости от конкретной задачи исследования. Если вес, придаваемый чувствительности методики, обозначить как w , то вес, придаваемый специфичности, составит $1 - w$. В этом случае оптимальный пороговый уровень, когда характеристики теста будут максимально удовлетворять требованиям исследователя, определяется максимальным значением M (Woodward M., 2014):

$$M = wSe + (1 - w) \cdot Sp.$$

Если цель состоит в том, чтобы добиться оптимальной классификации при равной весовой значимости чувствительности и специфичности, то $w = 0,5$. В этом случае уравнение имеет вид:

$$M = 0,5 \cdot Se + 0,5 \cdot Sp.$$

Для выявления оптимального порогового уровня диагностики можно использовать **индекс Юдена** (*Youden's index, J*), который рассчитывается следующим образом (Власов В. В., 2004; Szklo M., Nieto F. J., 2007):

$$J = Se + Sp - 1.$$

В рассматриваемом примере максимальные величины показателя M и индекса Юдена наблюдаются при диагностическом пороге 31 балл, который можно признать оптимальным для разделения испытуемых в зависимости от наличия/отсутствия распознаваемого заболевания, если предъявляются равные требования к специфичности и чувствительности теста (см. табл. 17).

Компромиссным решением может быть батарея тестов, когда на первом этапе применяются относительно доступные и достаточно чувствительные методики предварительной оценки вероятного риска распознаваемого состояния, а затем у лиц, характеризующихся повышенным риском, осуществляют более детальную диагностику, уже используя высокоспецифичные тесты. Иными словами, высокая чувствительность — это принцип скрининг-методик, а высокая специфичность — специальных диагностических методик.

Для интегральной оценки соотношения чувствительности и специфичности методики при различных пороговых уровнях диагностики часто используют так называемую **характеристическую кривую** (*receiver operating characteristic curve, ROC-curve*) (Armitage P. [et al.], 2001; Peat J. K., Barton B., 2014; Woodward M., 2014). Эта кривая отражает связь вероятностей ложноположительных (1-специфичность) и

истинно положительных (чувствительность) результатов. На рис. 43 представлена кривая, основанная на операционных характеристиках рассматриваемого примера гипотетического диагностического теста. Чем меньше изгиб кривой, чем ближе она расположена к биссектрисе, проходящей через начало координат под углом 45° , тем менее эффективно диагностическое исследование, так как точки на такой диагонали соответствуют отсутствию диагностической валидности теста (перемещение диагностического порога не приводит к изменению пропорции ложноположительных и истинно положительных результатов).

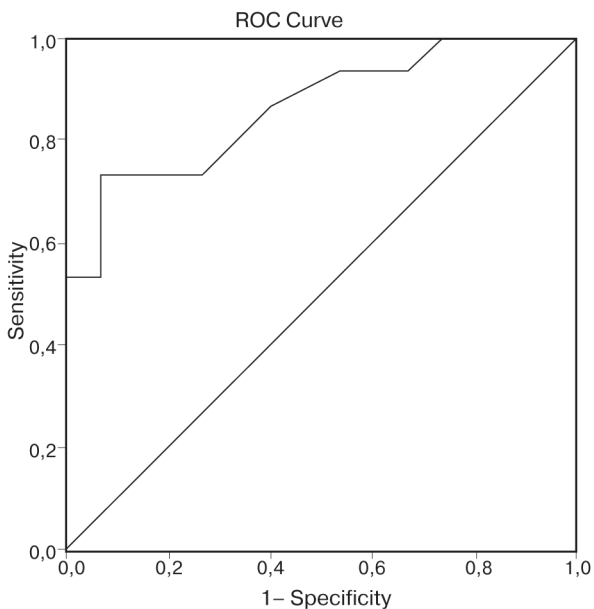


Рис. 43. Характеристическая кривая, основанная на данных рассматриваемого гипотетического диагностического теста (табл. 16)

Площадь под кривой (*area under the curve, AUC*), ограниченная характеристической кривой и осью абсцисс, является хорошей интегральной характеристикой эффективности теста. Теоретически эта площадь может принимать значения от 0 до 1,0. Поскольку в валидных тестах характеристическая кривая проходит выше биссектрисы, то обычно заслуживает обсуждения лишь изменение площади в диапазоне от 0,5 до 1,0 (площадь под биссектрисой изначально равна 0,5

и свидетельствует об отсутствии валидности методики). Чем ближе значение площади к 1,0, тем более эффективен диагностический тест. Статистические расчеты направлены на выявление достоверности различий вычисленной площади и площади, равной 0,5, свидетельствующей об отсутствии эффективности методики. В рассматриваемом примере площадь под кривой равна 0,864; данное значение с 99,9 % вероятностью ($P = 0,001$) отличается от неэффективного показателя, равного 0,5. Приведенные оценки свидетельствуют об эффективности методики. Отметим, что большинство прикладных статистических программ (SPSS, SAS, STATA) позволяют построить такие характеристические кривые и изучить их параметры.

Отношение правдоподобия при положительном результате теста говорит о том, во сколько раз вероятность положительного результата теста выше у пациента с действительным наличием заболевания по сравнению с пациентом, у которого заболевание отсутствует (Кельмансон И. А., 2002; Власов В. В., 2004; Armitage P. [et al.], 2001; Woodward M., 2014). Рассчитывается данный показатель следующим образом:

$$LR^+ = Se/(1 - Sp).$$

Отношение правдоподобия при отрицательном результате свидетельствует о том, во сколько раз вероятность отрицательного результата теста выше у пациента с наличием заболевания по сравнению с пациентом, у которого в действительности отсутствует заболевание (Кельмансон И. А., 2002; Власов В. В., 2004; Armitage P. [et al.], 2001; Woodward M., 2014). Данный показатель вычисляется следующим образом:

$$LR^- = (1 - Se)/Sp.$$

В приведенном примере диагностики симптомов постнатальной депрессии на основе показателей Эдинбургской шкалы (см. табл. 15) отношение правдоподобия составит:

- при положительном результате теста

$$LR^+ = Se/(1 - Sp) = 0,92/(1 - 0,88) = 7,67;$$

- при отрицательном результате теста:

$$LR^- = (1 - Se)/Sp = (1 - 0,92)/0,88 = 0,09.$$

Полученные цифры свидетельствуют о том, что при позитивных результатах теста вероятность наличия депрессии у женщины возрастает в 7,67 раза, при отрицательных результатах она возрастает в 0,09 раза (или уменьшается примерно в 11 раз).

Прогностичность (*predictive value*). Суждение о критериальной валидности диагностической методики во многом основывается на вероятности совпадения выявленных положительных результатов с истинным наличием заболевания и отрицательных результатов с его отсутствием. Для этих целей широко используются показатели прогностичности (Кельмансон И. А., 2002; Власов В. В., 2004; Rothman К. J. [et al.], 2008; Woodward М., 2014).

Прогностичность положительного результата (*predictive value positive, PV⁺*) отражает вероятность того, что пациент, имеющий положительные результаты теста, в действительности страдает заболеванием. Расчет осуществляется по формуле:

$$PV^+ = a/(a + b).$$

Прогностичность отрицательного результата (*predictive value negative, PV⁻*) отражает вероятность того, что у обследуемого с отрицательными результатами диагностического метода в действительности заболевание отсутствует. Расчет осуществляется по формуле:

$$PV^- = d/(c + d).$$

Пример

Применим данные об использовании Эдинбургской шкалы для диагностики симптомов депрессии (см. табл. 15). Прогностичность положительного результата составит:

$$PV^+ = a/(a + b) = 35/(35 + 26) = 0,57.$$

В 57 % случаев женщины, имеющие показатели по шкале депрессии, превышающие выбранный пороговый уровень, имеют клинически подтвержденные проявления постнатальной депрессии.

Прогностичность отрицательного результата составит:

$$PV^- = d/(c + d) = 186/(3 + 186) = 0,98.$$

В 98 % случаев отрицательные результаты теста позволяют отвергнуть диагноз «постнатальная депрессия».

Высокие показатели прогностичности отрицательного результата являются непременным условием любого диагностического метода, направленного на выявление редкого заболевания (состояния), так как, по условию, большая часть обследуемых лиц не страдает данным заболеванием.

Показатели прогностичности теста отчасти зависят от его чувствительности и специфичности. Действительно, чем чувствительнее тест, тем меньше вероятность того, что пациент, имеющий отрицательные диагностические результаты, страдает заболеванием и, сле-

довательно, тем выше прогностичность отрицательного результата. Аналогично, чем выше специфичность теста, тем меньше вероятность того, что у пациента с положительными диагностическими результатами отсутствует заболевание, т. е. тем выше прогностичность положительного результата. Однако нужно отметить, что при прочих равных условиях прогностичность диагностического исследования существенным образом зависит от преваленса (распространенности в популяции) распознаваемого заболевания (патологического состояния), что не учитывается при расчете показателей чувствительности и специфичности. Это положение особенно актуально в случаях диагностики относительно редких заболеваний, с чем сталкиваются исследователи при проведении скрининга. Преваленс заболевания (патологического состояния) в обследуемой популяции и операционные характеристики теста, взаимодействуя между собой, определяют, какова вероятность идентификации заболевания среди обследуемых лиц. Лучше всего указанная взаимосвязь выявляется, если вместо преваленса использовать производную от него величину — *претестовые шансы заболевания (pretest odds)* (Кельмансон И. А., 2002; Straus S. E., 2011). Эти показатели соотносятся следующим образом:

Претестовые шансы заболевания = преваленс / (1 – преваленс).

Тогда **послетестовые шансы заболевания (post-test odds)** будут вычисляться как произведение претестовых шансов и отношения правдоподобия при положительных результатах теста:

Послетестовые шансы = претестовые шансы $\times LR^+$.

Конструктивная валидность (construct validity) оценивается с целью определить, насколько структура опросника позволяет достоверно измерить то, что он должен измерить. Строятся определенные гипотезы, в основе которых лежат различные факторы (например, психологические, социальные или клинические показатели). В процессе исследования эти теоретические предположения подтверждают или опровергают. Различают виды конструктивной валидности:

— *конвергентная (convergent validity)*: результаты двух способов измерения одной характеристики должны коррелировать между собой;

— *дискриминантная (discriminant validity)*: результаты измерения различных характеристик не связаны между собой (Barker C. [et al.], 2002).

В обоих случаях конструктивная валидность предполагает оценку взаимосвязи изучаемой шкалы с определенными характеристиками.

Факторный анализ по методу главных компонент (Howell D. C., 2013) — статистический прием, который обеспечивает

проверку предполагаемой структуры опросника и призван выявить отдельные компоненты в пределах структуры опросника. Относится к общепринятым подходам к оценке конструктивной валидности. В процессе проведения факторного анализа определяют то, как пункты опросника распределяются по отношению к своим шкалам. Кроме того, факторный анализ должен показывать, что пункты шкалы относятся к своему фактору и тесно с ним коррелируют, слабо коррелируя с другими факторами.

Пример

Опросник SF-36, предназначенный для оценки качества жизни, связанного со здоровьем, принципиально предполагает наличие двухкомпонентной структуры (физического и психологического компонентов здоровья). Пункты таких шкал «Физическое функционирование» и «Боль» имеют высокую «факторную нагрузку» на физический компонент здоровья и низкую — на психологический компонент (Новик А. А., Ионова Т. И., 2007).

Метод известных групп. Респондентов распределяют на группы в зависимости от наличия или отсутствия у них какого-либо фактора. Высказывается наиболее вероятное предположение в отношении распределения у них данного фактора, и проводится анализ взаимосвязей показателей в зависимости от изучаемого фактора.

Пример

Изучаются показатели качества жизни, связанного со здоровьем, в зависимости от возраста пациентов: высказывались предположения о различиях показателей физического здоровья в различных возрастных группах. Эти предположения были подтверждены: респонденты старших возрастных групп имели показатели качества жизни хуже, чем респонденты молодого возраста (Новик А. А., Ионова Т. И., 2007).

Еще одним способом оценки конструктивной валидности является оценка корреляций с внешними критериями (шкалами другого опросника, симптомами, клиническими или лабораторными показателями и т. д.). В таком случае целесообразно использовать **многочертную-многометодную матрицу** (*multitrait-multimethod matrix*) (Barker С. [et al.], 2002). Данная методика была предложена исследователями Д. Кэмпбллом и Д. Фиске (D. Campbell, D. Fiske). Матрица представляет собой таблицу, отражающую корреляции между различными методиками, измеряющими различные конструкты. Важно одновременно анализировать и соотносить друг с другом не только измеряемые черты (т. е. что измеряется), но и измерительные методы (т. е. как измеряется).

Пример

Исследуется конструкт «высокий уровень тревоги во время публичных выступлений». Можно изучить показатели тревоги при помощи, скажем, двух различных опросников, а также исследовать частоту сердечных сокращений, кожный гальванический рефлекс у испытуемых во время публичных выступлений. Кроме того, у испытуемых можно исследовать коэффициент интеллекта, выраженность экстраверсии, уровень самооценки.

Многочертно-многометодная матрица будет отражать взаимную корреляцию всех изученных показателей. В частности, матрица будет показывать выраженность положительных корреляций измерений, относящихся к изучаемому конструкту, с измерениями, относящимися к другим близким конструктам. Это будет **конвергентная валидность**. С другой стороны, матрица будет показывать слабость или отсутствие корреляций измерений, относящихся к изучаемому конструкту, с измерениями, относящимися к конструктам, которые не связаны с данным конструктом (**дискриминантная валидность**). В приведенном примере все измерения, направленные на изучение тревоги во время публичных выступлений, должны хотя бы умеренно коррелировать друг с другом. Предположительно, они не должны достоверно коррелировать с возрастом и коэффициентом интеллекта испытуемых. Кроме того, они должны лишь незначительно коррелировать с показателем самооценки, однако гораздо сильнее коррелировать с выраженностью экстраверсии. Матрица также позволяет выявить степень вариабельности отдельных методик, тенденцию, с которой коррелируют между собой близкие методики. Например, различные опросники, основанные на самоотчетах, часто взаимно коррелируют, хотя и призваны оценивать различные конструкты. Вот почему важно использовать в ходе исследования различные методики, а не целиком полагаться на какой-то один тип измерения.

3.2.3. Альтернативные подходы к оценке качества измерений

Наряду с традиционными подходами, основанными на исследовании надежности и валидности измерений, существует ряд альтернативных подходов.

Расширенная статистическая теория (*generalizability theory, G-theory*) (Cronbach L. J., 1972). Классическая и расширенная статистические теории измерений имеют общий предмет рассмотрения — оценку истинных и ошибочных компонент измерения с целью определить надежность тестовых результатов. Обе теории применимы к обработке данных на выборках небольшого объема. Расширенная статистическая теория позволяет вести поиск нескольких источников

систематических и случайных погрешностей измерения. Если в классической теории предполагается, что дисперсии истинных и ошибочных компонентов измерения, а также корреляции с внешним критерием в параллельных вариантах тестирования являются одинаковыми, то в расширенной статистической теории параллельные варианты теста считаются случайными выборками из одной генеральной совокупности заданий. В классической теории ошибочный компонент измерений рассматривается как случайная погрешность неизвестного происхождения, что не позволяет в ее рамках исследовать надежность результатов опроса в зависимости от таких источников погрешностей, как нестабильность результатов испытуемых (низкий показатель надежности при повторном тестировании), недостаточное внутреннее согласие, несогласованность экспертов (низкий уровень межисследовательского согласия). Перечисленное может быть оценено с помощью аппарата расширенной статистической теории. Расширенная статистическая теория основана на многофакторной оценке и предполагает ответы на вопросы:

— Для каких наблюдений возможно обобщение, основанное на имеющихся конкретных наблюдениях? или

— В отношении каких ситуаций те измерения, которые были получены на данных конкретных наблюдениях, могут считаться репрезентативными?

При подобном подходе меньше внимания уделяется точности измеренных показателей, и больше внимания — источникам вариации измеренных показателей. Разграничение надежности и валидности измерений становится достаточно условным. Предполагается, что измерение и, соответственно, ошибка измерения включают в себя конкретных субъектов измерения, измеряемые переменные и так называемые фасы (*facets*) (условия измерения). Можно выделить четыре принципиальных условия: наблюдателей, ситуации, инструменты и обстоятельства. С каждым из них связана определенная ошибка измерения, и задача состоит в том, чтобы ее идентифицировать. Указанные условия соотносятся с традиционными психометрическими понятиями (табл. 18). Допустим, два независимых исследователя изучают признаки, характеризующие поведение детей, и дважды проводят оценку. Может ли оценка, даваемая каждым отдельным исследователем в любое время, рассматриваться как некая универсальная оценка, которую можно сопоставлять с оценками других исследователей, данным в другое время, отражая, таким образом, некоторую универсальную оценку? Исследователь и время оценки в данном примере выступают как условия универсального измерения. В рамках расширенной статистической теории дизайн такого исследователя будет выглядеть так:

$$p \times r \times o,$$

где p — наблюдаемый ребенок; r — наблюдатель; o — момент наблюдения.

Таблица 18

Связь надежности и валидности методики с различными фасетами
(Cronbach L. J., 1972; Barker C. [et al.], 2002)

Фасет	Соответствующее психометрическое понятие
Наблюдатели: различия у отдельных исследователей	Надежность методики по показателю межисследовательского согласия
Ситуации: различия на отдельных временных этапах	Надежность методики при повторном тестировании, предсказательная валидность методики
Инструменты: различия в способах измерения одного и того же показателя (включая различия отдельных признаков)	Эквивалентные формы, внутреннее согласие, конкурентная валидность, конвергентная валидность
Ситуации (обычно переход от более контролируемых к менее контролируемым ситуациям)	Критериальная валидность, конвергентная валидность

Показатель, наблюдаемый у ребенка p наблюдателем r в момент времени o , можно обозначить как X_{pro} . Вариация показателей будет описываться следующим уравнением (Shavelson R. J., Webb N. M., 1991):

$$\sigma^2(X_{pro}) = \sigma_p^2 + \sigma_r^2 + \sigma_o^2 + \sigma_{pr}^2 + \sigma_{po}^2 + \sigma_{ro}^2 + \sigma_{pro,e}^2.$$

Таким образом, универсальная вариация признака $\sigma^2(X_{pro})$ может быть разделена на независимые источники вариации:

- вариацию, связанную с различиями между отдельными детьми σ_p^2 ;
- вариацию, связанную с различиями между отдельными наблюдателями σ_r^2 ;
- вариацию, связанную с отдельными временными этапами проведения измерения σ_o^2 ;
- вариацию, связанную со взаимодействием «наблюдатель — ребенок», т. е. отражающую разную оценку, даваемую отдельными наблюдателями детям σ_{pr}^2 ;
- вариацию, связанную со взаимодействием «ребенок — момент наблюдения» σ_{po}^2 , отражающую различия поведения детей в отдельные моменты;

— вариацию, связанную со взаимодействием «исследователь — момент наблюдения» σ^2_{ro} , отражающую разность восприятия наблюдателями поведения детей в отдельные моменты наблюдения.

Наконец, резидуальную (остаточную) вариацию) $\sigma^2_{pro,e}$, которая отражает взаимодействие «ребенок — наблюдатель — время», а также вариацию, связанную со случайной ошибкой e . Источник этой ошибки может быть связан с неучтенными условиями исследования (например, в момент наблюдения ребенок плохо себя чувствует).

Основная задача — вычислить долю вариации, приходящуюся на различия между отдельными детьми в общей вариации, и сопоставить ее с долями вариации, приходящимися на другие ее источники. Расчеты во многом близки к тем, которые используются при вычислении коэффициента внутриклассовой корреляции. Чаще всего анализ указанной модели строится по методике дисперсионного (вариационного) анализа (ANOVA) и может быть осуществлен при помощи большинства прикладных компьютерных программ (SPSS, SAS, STATA). Имеются и специальные доступные статистические программы, предназначенные для реализации расширенной статистической теории (например, EduG) (<http://www.irdp.ch/edumetrie/englishprogram.htm>) (Cardinet J. [et al.], 2010).

Метрическая система Раша (*item response theory*) (DeMars С., 2010) лучше всего реализуется при изучении качества тестов, направленных на изучение способностей, знаний, навыков исследуемых, хотя возможны и иные сферы применения. Основной принцип состоит в том, чтобы сопоставить вероятность правильного ответа на вопрос или выполнения задания (чаще с альтернативной оценкой: да/нет, выполнил/не выполнил) с тем конструктом, который тест призван оценить. Этот конструкт обычно не измеряется непосредственно, а лишь предполагается, поэтому именуется латентным. Характеристическая кривая признака (*item-characteristic curve*), отражающая связь этих показателей, чаще имеет S-образную удлиненную форму. Чаще всего для выявления связи между изучаемым латентным показателем и вероятностью правильного ответа на вопрос используется модель, предложенная датским статистиком Георгом Рашем (G. Rasch). Он разработал метод предсказания вероятности того, что испытуемый правильно ответит на вопрос бинарного типа «верно/неверно» на основе процента в генеральной совокупности других респондентов, ответивших на этот вопрос правильно, и процента других вопросов, на которые данный испытуемый уже ответил правильно. Первый показатель отражает уровень трудности вопроса (*item difficulty*, β_i), второй — способности испытуемого (*proficiency*, θ). Выполнение тестов человеком, не знакомым с предлагаемым набором вопросов, можно предсказать с погрешностью, поддающейся

вычислению, даже если давать испытуемым разные тесты. Модель Раша была использована при оценке возможности использования опросника Five-to-Fifteen, предназначенного для выявления отклонений поведения у детей от 5 до 15 лет на основании оценок родителей (Illum N. O., Gradel K. O., 2014).

Вопросы для самоконтроля

1. Что такое методы и методики измерения?
2. Каковы важнейшие методы измерений, используемые в клинической психологии? В чем их достоинства и недостатки?
3. Что такое надежность измерения? Каковы подходы к ее оценке?
4. Что такое валидность измерения? Каковы важнейшие виды валидности измерения?

АНАЛИЗ ДАННЫХ

4.1. Представление и суммирование данных

Клинико-психологические исследования предполагают сбор информации о многочисленных характеристиках, отражающих состояние здоровья большого числа обследуемых лиц. Очевидно, представление этих сведений в их исходном виде невозможно, да и лишено смысла. С целью представить данные содержательно и вместе с тем в компактной форме, обеспечивающей сопоставимость важнейших характеристик обследованных, используются методы описательной статистики (*descriptive statistics*). При этом часть первичной информации неизбежно теряется (редукция данных), что диктует необходимость понимания важнейших принципов, используемых в отдельных описательных статистических подходах, и выбора оптимальной описательной статистики в зависимости от конкретной задачи (Кельмансон И. А., 2002; Howell D. C., 2013).

Предпосылкой верного использования принципов описательной статистики является понимание того, какого рода переменные (*variables*) используются в исследованиях.

Дискретные переменные (*discrete*) принимают значения, укладываемые в ограниченное число категорий; они не могут принимать промежуточные значения. *Дихотомная* дискретная переменная) может принимать альтернативные значения по принципу «или-или» (пол, возникновение заболевания, воздействие какого-либо фактора и т. п.). *Мультихотомная* дискретная переменная может принимать больше двух значений, примерами могут служить группа крови, раса, и т. п.

Номинальная шкала значений не характеризуется каким-то внутренним порядком возрастания или убывания значений, как в примерах выше.

Ординальная шкала используется, когда возможные значения дискретных переменных имеют естественную прогрессию. Например, к числу ординальных дискретных переменных можно отнести степень недостаточности кровообращения, степень активности ревматизма, стадию костных изменений при ревматоидном артрите, степень интеллектуальных нарушений и ряд других.

Численно дискретные переменные имеют количественный и целочисленный характер, например число предшествовавших родов, беременностей, аборт у матери, число приступов бронхиальной астмы за год и т. п.

Непрерывные переменные могут теоретически принимать любые численные значения в пределах заданного диапазона. К ним отнесены многие клинические параметры, например масса, длина тела, уровень гемоглобина, величина артериального давления.

Подразделение переменных на дискретные и непрерывные имеет существенное значение с точки зрения оптимального выбора способа представления и анализа данных исследований. Переменные, относящиеся к указанным двум различным классам, дают ответы и на различные вопросы, связанные с анализом данных. Дискретные переменные позволяют сопоставить долю лиц, относящихся к отдельным категориям, а непрерывные переменные дают возможность сопоставить средние значения.

Некоторые переменные могут быть классифицированы лишь единственным возможным способом, что связано с их характером. Однако существуют ситуации, в которых способ классификации переменных зависит от выбора исследователя. Например, возраст ребенка может быть оценен в исходной форме как непрерывная переменная или в зависимости от заданных возрастных диапазонов (ординальная дискретная величина), причем в последнем случае возрастные диапазоны могут быть сформированы произвольно, и выбор может быть продиктован конкретными задачами исследования.

Пример

В случае исследования связи возраста матери с риском рождения ребенка с синдромом Дауна представляется разумным классифицировать матерей в зависимости от возраста в виде следующих ординальных категорий: «моложе 30 лет», «30–34 года», «35–39 лет», «40 лет и старше». Такое подразделение оправдано в связи с тем, что риск синдрома Дауна чаще наблюдается у детей, родившихся у матерей старше 30 лет, и предложенная группировка наблюдений учитывает данное обстоятельство, объединяя женщин моложе 30 лет в одну категорию. В то же время приведенная возрастная группировка оказалась бы неэффективной в случае анализа ассоциации риска осложненного течения беременности (например, возникновения аффективных расстройств) с возрастом матери, так как риск данного профиля патологии часто наблюдается у подростков, и включение в единую группу женщин моложе 30 лет может замаскировать имеющуюся ассоциацию.

Укрупнение диапазонов переменных, приводящее к редуцированию первичных данных, будучи привлекательным с точки зрения более компактной и доступной для восприятия формы представления результатов, таит в себе опасность невозможной утраты части ин-

формации. Оправданным можно считать сбор детальной информации на начальных стадиях исследования с последующим ее редуцированием (в случае необходимости), нежели сбор исходных данных в редуцированной форме.

Не существует общепринятых рекомендаций относительно того, каковы критерии выбора тех или иных диапазонов значений переменной, переводимой в категориальную форму. Нередко выбор носит интуитивный характер. Возможны также ситуации, в которых первичная категоризация значений выявляет незначительные различия между двумя соседними категориями на ординальной шкале, что позволяет укрупнить диапазоны, объединив несколько соседних категорий.

Адекватное представление результатов исследования предполагает наличие сведений о распределении изученных переменных. Для дискретных переменных наиболее информативной суммарной статистикой является частотное распределение наблюдений в пределах отдельных категорий. Чаще всего используются таблицы, отражающие распределение частот (табл. 19), столбчатые (рис. 44) и круговые (рис. 45) диаграммы.

Таблица 19

**Пример таблицы,
отражающей распределение категорий признака**

Образование матери	Число наблюдений	%
Неполное среднее	5	2,4
Среднее	115	54,7
Среднее специальное	25	11,9
Незаконченное высшее	10	4,8
Высшее	55	26,2
Суммарно	210	100

Как представить результаты

— При указании процентов следует приводить лишь один десятичный знак и лишь в тех случаях, когда число наблюдений превышает 100.

— Если число наблюдений менее 100, проценты следует указывать без десятичных знаков.

— Когда общее число наблюдений менее 20, на проценты вообще не нужно ссылаться; рекомендуется указывать лишь абсолютные значения, относящиеся к отдельным категориям признака (например, из 18 обследованных детей 10 были практически здоровы).

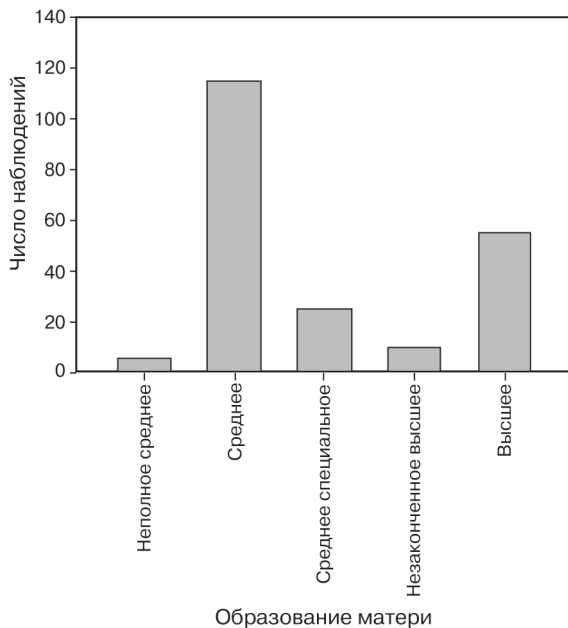


Рис. 44. Пример столбчатой диаграммы («bar chart»), отражающей распределение категорий признака



Рис. 45. Пример круговой диаграммы («pie chart»), отражающей распределение категорий признака

— Если общее число наблюдений меньше 10, рекомендуется числа выражать словами, а не цифрами (например, три из восьми обследованных пациентов имели симптомы депрессии) (Peat J. K., 2002; Peat J. K., Barton B., 2014).

Суммарная статистика для переменных с непрерывным распределением требует как минимум двух параметров. Первый параметр должен отражать **меру центральной тенденции** (локализации-наблюдений), а второй — **меру вариации** (разброса наблюдений). Информация о них требуется для описания формы распределения наблюдений. Рис. 46 иллюстрирует распределения A и B, которые имеют одинаковую центральную тенденцию (центры обоих распределений совпадают по оси абсцисс, но вариация распределения B (разброс наблюдений) больше, чем распределения A, где имеющиеся наблюдения более «компактно» концентрируются вокруг центрального значения. Рис. 47, напротив, иллюстрирует распределения A и B сходной формы, имеющие одинаковую вариацию значений признака, но локализация этих распределений (центральная тенденция) различна. Очевидно, изолированный учет одного из двух характеристических параметров распределений не дал бы полного представления об их особенностях.

Для оценки центральной тенденции чаще всего используются три показателя: среднее, медиана и мода. Каждая характеристика, направленная на выявление наиболее типичных значений непрерывно распределенной переменной в исследуемой популяции, имеет определенные преимущества и недостатки.

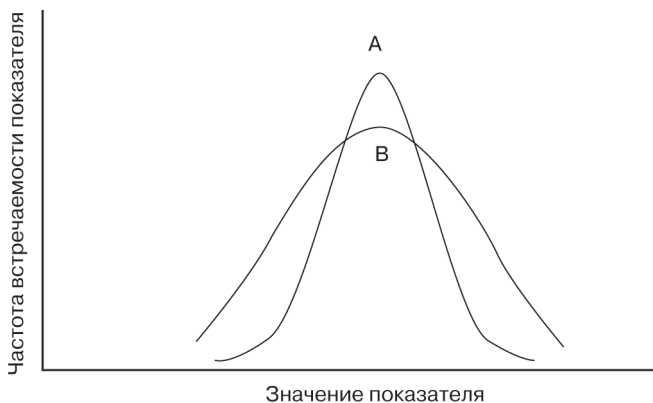


Рис. 46. Пример двух распределений с одинаковой центральной тенденцией и различными вариациями показателей

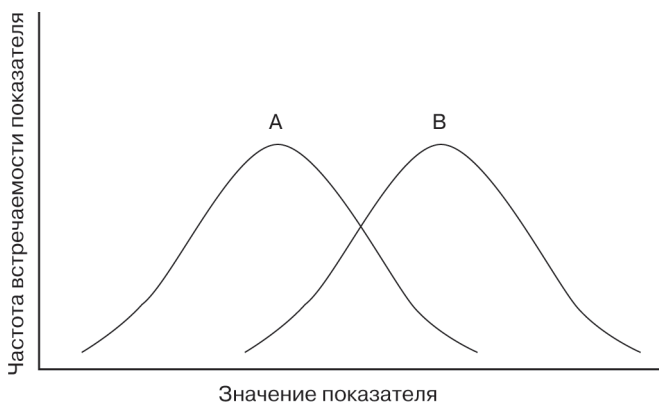


Рис. 47. Пример двух распределений с различными центральными тенденциями и одинаковыми вариациями показателей

Среднее значение чаще всего используется в качестве характеристики центральной тенденции. Оно вычисляется как простое среднее арифметическое:

$$\bar{x} = \frac{\sum x}{n},$$

где $\sum x$ — сумма всех значений; n — число наблюдений.

Показатель оптимален для описания наблюдений, подчиняющихся закону нормального распределения (Howell D. C., 2013; Peat J. K., Barton V., 2014). Однако среднее весьма чувствительно к крайним значениям распределения, на него могут существенно повлиять единичные экстремальные показатели.

Пример

Предположим, что у 5 обследованных новорожденных детей значения массы тела составили: 2900, 3000, 3000, 3100 и 4500 г. Среднее значение массы тела составит: $(2900 + 3000 + 3000 + 3100 + 4500)/5 = 16500/5 = 3300$ г. Очевидно, средняя величина, смысл которой заключается в отражении типичного значения для данного распределения, превышает реальные значения массы тела четырех из пяти изученных наблюдений, а следовательно, вряд ли может рассматриваться в качестве удовлетворительной характеристики распределения. Пример является вариантом такого распределения, при котором численность наблюдений, приходящихся на различные стороны распределе-

ния, существенно различается. В таких случаях говорят, что распределение характеризуется асимметрией, смещено (*skewed*). В данном примере большая часть наблюдений смещена в область малых значений, влево (отрицательная асимметрия распределения). Следовательно, чем больше выражена асимметрия распределения, тем менее пригодным становится среднее значение для описания центральной тенденции.

Медиана, или 50-й центиль, в буквальном смысле соответствует середине распределения. Значение является медианой распределения в том случае, если половина наблюдений имеет величины рассматриваемого показателя ниже этого значения, а половина — выше. Для определения медианы распределения все наблюдения условно располагаются в порядке возрастания или убывания. Если число наблюдений нечетное, позиция наблюдения, соответствующего медиане, определяется как $(n + 1)/2$.

Пример

Если в выборке имеется 49 наблюдений, то медиане будет соответствовать значение 25-го наблюдения в упорядоченном ряду. Если число наблюдений четное, медиане будет соответствовать среднее арифметическое значений, соответствующих наблюдениям, находящимся на позиции $n/2$ и $(n/2 + 1)$.

Преимуществом использования медианы в качестве меры центральной тенденции является то, что она не подвержена влиянию экстремальных значений распределения. Так, в примере с массой тела новорожденных детей медиане будет соответствовать показатель 3000 г, который значительно лучше, чем среднее, отражает типичное значение данного распределения. Вот почему при описании центральной тенденции выборки, не подчиняющейся законам нормального распределения, рекомендуется использовать именно медиану, а не среднее значение (Howell D. C., 2013; Peat J. K., Barton B., 2014). Это особенно актуально в случае выборки малой численности.

Недостатки:

- Медиана, отражая лишь ранговое значение, не дает информации об остальных значениях распределения.
- Значение медианы распределения труднее использовать для изучения статистической достоверности различий двух распределений.

Мода распределения — наиболее часто встречающееся значение в данной выборке. Мода редко используется в качестве единственной описательной характеристики центральной тенденции.

Преимущество

- Принципиально важным для получения дополнительной информации может оказаться выявление двух или нескольких мод рас-

пределения (би- и полимодальности), что может явиться аргументом в пользу гетерогенности (неоднородности) распределения. Однако следует отметить, что анализ на предмет наличия полимодальности распределения должен осуществляться с особой тщательностью, так как нередко асимметричность распределения (особенно положительная, т. е. со сдвигом в сторону высоких значений, вправо) может неверно интерпретироваться как второй пик, что чревато ложными выводами.

Недостатки:

— При небольшом числе наблюдений высока вероятность того, что каждое значение может встретиться всего лишь один раз, и, следовательно, моды вообще не будет как таковой.

— Мода еще меньше, чем медиана, приемлема для статистических расчетов.

Мера вариации показателя — расположение значений, соответствующих отдельным наблюдениям, относительно центрального значения, представляет собой второй важный элемент описательной статистики. Простейшим показателем вариации можно считать **вариационный размах** (*range*), который вычисляется как разность между максимальным и минимальным значениями признака. В частности, в приведенном примере массы тела у новорожденных детей вариационный размах составит: $4500 - 2900 = 1600$ г.

Преимущество

Показатель вариационного размаха легко рассчитывается и интерпретируется.

Недостатки:

— Вариационный размах не является оптимальным в оценке вариации.

— Вариационный размах не отличается стабильностью: он может возрастать по мере увеличения численности обследуемой группы.

— Он не может быть непосредственно использован для статистических расчетов и тестов.

— Вариационный размах учитывает лишь крайние значения: в тех случаях, когда большая часть значений распределения концентрируется вокруг центрального, вариационный размах может дать неверное впечатление о характере распределения.

Более предпочтительной является такая мера вариации, которая позволила бы учитывать все отдельные значения показателей. Наиболее информативной и чаще всего используемой характеристикой, удовлетворяющей поставленному условию, является **вариация** (*variance*) и связанное с ней **стандартное отклонение (СО)** σ (*standard deviation, SD*). Каждый из этих показателей по-разному отражает степень дисперсии индивидуальных значений вокруг среднего. Для

того чтобы определить вариацию Var , сначала необходимо вычислить разницу между значением каждого наблюдения x и средним значением \bar{x} . Затем полученная разность возводится в квадрат ($[x - \bar{x}]^2$), благодаря чему отрицательные и положительные величины разности не уничтожают друг друга, а затем полученные значения суммируются ($\sum [x - \bar{x}]^2$). Сумма квадратов отклонений делится на число наблюдений минус 1 ($n - 1$). Таким образом, формула, используемая для вычисления вариации, имеет следующий вид:

$$Var = \frac{(\sum [x - \bar{x}]^2)}{(n - 1)},$$

или (то же самое):

$$Var = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{(n - 1)}.$$

В приведенных расчетах вариации сумма квадратов отклонений от среднего делится на $n - 1$, а не на n ; это позволяет осуществлять коррекцию с учетом того, что обследуется лишь выборка из генеральной совокупности, а не вся генеральная совокупность. Если расчеты касаются популяции в целом, в знаменателе допустимо использовать значение n .

СО равно квадратному корню вариации:

$$\sigma = \sqrt{Var}.$$

Чаще всего представление результатов описательной статистики непрерывных значений осуществляется так: среднее значение $\pm \sigma$. Если распределение значений не отличается от нормального распределения, эти два параметра дают его исчерпывающее описание. Примерно 68 % наблюдений, соответствующих нормальному распределению, будут находиться в диапазоне среднее $\pm 1 \sigma$, приблизительно 95 % — в диапазоне среднее $\pm 2 \sigma$, и около 99 % — в диапазоне среднее $\pm 2,5 \sigma$.

В тех случаях, когда требуется описание распределения, отличного от нормального, лучше описать его, используя пять показателей: указать значение медианы распределения и вариационного размаха или **межквартильного размаха** (*interquartile range*). **Квартили** — это значения случайной переменной, делящие распределение на четыре равные по объему части. Таким образом, первый квартиль рас-

пределения Q_1 — значение переменной, величины меньше которого имеют 25 % наблюдений. Второй квартиль Q_2 — значение, величины меньше которого имеют 50 % наблюдений (то же самое, что медиана). Наконец, третий квартиль Q_3 — значение, величины меньше которого имеют 75 % наблюдений.

Пример

Имеется девять значений признака: 15, 3, 9, 3, 14, 20, 7, 8, 11. Для того чтобы определить вариационный размах, медиану и квартили распределения, расположим значения в порядке возрастания: 3, 3, 7, 8, 9, 11, 14, 15, 20. Минимальное значение равно 3, максимальное равно 20 (вариационный размах). Посередине располагается значение, равное 9, — четыре наблюдения имеют меньшие значения, четыре большие. Это медиана, или второй квартиль. Значение первого квартиля равно 7, так как меньшие значения имеют два наблюдения, а большие — шесть наблюдений, что дает соотношение 1 : 3. А третьему квартилю соответствует значение, равное 14 (шесть наблюдений имеют значения меньше данного и два — больше, соотношение 3 : 1).

Приведен пример табличного представления показателей (табл. 20).

Таблица 20

Пример представления показателей для непрерывных переменных

Показатель	Число наблюдений n	Среднее (СО) или медиана (межквартильный размах)
Масса ребенка при рождении, г	139	2464,0 (514,6)
Гестационный возраст, нед.	133	36,6 (2,0)
Длительность пребывания ребенка в стационаре, день	132	27 (15–32)

Иногда представляет интерес сопоставление степени вариации изучаемого фактора в двух популяциях или двух переменных в одной изучаемой популяции. Например, в какой возрастной группе: среди детей раннего школьного возраста или у подростков — показатель уровня тревоги, измеренный по стандартной методике, подвержен большей вариабельности. Можно также задаться целью сопоставить вариабельность показателей тревожности и гиперактивности у детей раннего школьного возраста. СО является мерой дисперсии значений вокруг среднего, и для ответа на поставленные вопросы не представляется возможным непосредственно сопоставлять значения σ , так как при одном и том же значении σ вариабельность показателей будет тем большей, чем меньше среднее значение.

Коэффициент вариации (*coefficient of variation*), который определяется как отношение CO к среднему значению, позволяет учесть указанное выше обстоятельство. Часто коэффициент вариации K выражается в процентах:

$$K = \frac{(CO)}{x} \cdot 100 \%$$

Коэффициенты вариации могут сопоставляться непосредственно. **Стандартная ошибка среднего (СОС)** (*standard error of the mean, SEM*) — мера вариабельности, которая нередко используется в исследованиях. Хотя значение СОС непосредственно связано со значением CO , эти характеристики имеют принципиально различный смысл. CO характеризует степень разброса отдельных значений изучаемого параметра в выборке вокруг среднего значения; СОС определяет вариабельность среднего значения в выборке как функцию истинного среднего значения показателя в генеральной совокупности, из которой была сформирована данная выборка.

Пример

Поставлена задача определить: чему равно среднее значение показателя тревоги, оцененного на основании опросника Ахенбаха, у детей младшего школьного возраста в Санкт-Петербурге. Можно обследовать выборку из 1000 детей младшего школьного возраста и определить у них среднее значение показателя. Если обследовать других 1000 детей, трудно предположить, что вычисленное среднее значение в этой выборке будет совпадать со значением, полученным в первой выборке. Однако обе выборки будут давать оценку среднего популяционного значения показателей тревоги, и можно предположить, что среднее значение, полученное из средних значений в этих двух выборках, будет больше соответствовать истинному среднему популяционному значению, чем каждое среднее значение в отдельности. Если продолжать формирование все новых и новых выборок, будет получена серия средних значений, и при достаточно большом числе этих выборок их средние значения сами по себе уже будут формировать нормальное распределение показателей независимо от первичных индивидуальных значений показателей, соответствующих отдельным наблюдениям в каждой выборке. Среднее значение из таких средних для выборок значений будет максимально соответствовать популяционному среднему.

Подобно тому как вычисляется вариабельность индивидуальных значений показателя в выборке относительно выборочного среднего, можно оценить стандартную ошибку среднего — вариабельность средних выборочных значений относительно истинного популяционного среднего. Она говорит о том, в какой мере выборочное среднее значение соответствует истинному популяционному среднему значе-

нию. Очевидно, точность, с которой можно в выборочном исследовании оценить истинное среднее популяционное значение, зависит от вариабельности значений в выборке и от численности выборки.

Формула для расчета СОС выглядит следующим образом:

$$\text{СОС} = \frac{(\text{СО})}{\sqrt{n}}.$$

Таким образом, СОС позволяет оценить интервал, в пределах которого с заданной вероятностью находится истинное среднее популяционное значение изучаемого показателя, и сопоставить средние значения в различных популяциях. Следовательно, важным моментом при описании данных и их дальнейшей статистической обработке является проверка того, подчиняется ли изучаемый непрерывный показатель нормальному распределению. Возможные подходы к проверке нормальности распределения:

- сопоставление значений среднего и медианы: чем ближе эти значения, тем больше вероятность того, что распределение имеет характер нормального (простой, хотя и нестрогий подход);

- в случае нормального распределения 95 % всех показателей находятся в диапазоне: среднее значение $\pm 1,96\sigma$.

Иными словами, большая часть значений показателя должна находиться в диапазоне: среднее минус два среднеквадратических отклонения – среднее плюс два среднеквадратических отклонения. Таким образом, простым приемом для проверки распределения на предмет его соответствия нормальному распределению является следующий: умножить рассчитанное СО на два, вычесть данное значение из среднего и прибавить данное значение к среднему. Полученные величины должны примерно соответствовать минимальному и максимальному значениям показателей в выборке. Если два рассчитанных показателя ниже минимального и максимального значений, то распределение смещено вправо (положительная асимметрия); если существенно выше минимального и максимального, то распределение характеризуется отрицательной асимметрией. Таким образом, в качестве рабочего правила можно принять следующее: в случае нормального распределения величина среднеквадратического отклонения не может превышать половины величины среднего значения. В противном случае рассчитанное минимальное значение в выборке будет отрицательным, что невозможно (Peat J. K., Barton B., 2014).

Формальным подходом к оценке нормальности распределения является использование статистических критериев: Колмогорова – Смирнова или Шапиро – Уилка (Peat J. K., Barton B., 2014). Большинство прикладных статистических программ позволяют применить эти критерии.

4.2. Оценка влияния случая на результаты исследования

4.2.1. Проверка гипотезы

Оценка обоснованности выдвинутой гипотезы, прежде всего, предполагает оценку значимости полученных результатов. Нужно решить, в какой мере выявленные ассоциации между патогенным воздействием и эффектом могут быть связаны с какими-то альтернативными объяснениями. Если представляется возможным отвергнуть подобные альтернативные объяснения, можно сделать вывод о внутренней валидности (*internal validity*), значимости полученных результатов (Barker C. [et al.], 2002; Rothman K. J. [et al.], 2008; Woodward M., 2014). В противном случае можно предположить, что полученные результаты являются следствием случайного совпадения.

Оценка возможного влияния случая на результаты исследования состоит из двух взаимосвязанных компонентов. Первый заключается в выполнении теста, направленного на определение статистической достоверности того, что полученные результаты в действительности отражают картину в генеральной совокупности, а не связаны с вариабельностью показателей в отдельных выборках. Второй компонент состоит в оценке доверительных интервалов и определении тех диапазонов, в пределах которых с заданным уровнем вероятности находятся истинные значения, отражающие величину эффекта. Теоретической основой указанных подходов является так называемый принцип инференции (*inference*), суть которого состоит в попытке делать выводы о более широкой группе индивидуумов, опираясь на данные, полученные в более узкой группе.

Нулевая гипотеза H_0 . Проведение теста, направленного на оценку статистической значимости выявленной ассоциации, предполагает четкую формулировку нулевой гипотезы H_0 , которая нуждается в опровержении. Нулевая гипотеза является допущением того, что ассоциация между воздействием и изучаемым эффектом в действительности отсутствует. Альтернативной гипотезой, которую можно обозначить как H_1 или H_A , будет предположение о существовании реальной связи между воздействием и эффектом.

Пример

Если в когортном исследовании пропорция лиц в контрольной группе наблюдений, у которых развилось заболевание, составляет p_0 , а аналогичная пропорция в группе, подвергшейся воздействию, составляет p_1 , нулевая гипотеза, согласно которой связь между воздействием и риском заболевания отсутствует, предполагает равенство пропорций в двух сопоставляемых группах. Сказанное имеет следующее математическое выражение:

$$H_0: p_0 = p_1.$$

Альтернативная гипотеза о существовании реальной ассоциации между воздействием и риском развития заболевания будет иметь выражение:

$$H_1: p_0 \neq p_1.$$

После того как сформулированы альтернативные гипотезы H_0 и H_1 , представляется возможным выполнение теста на предмет оценки статистической достоверности ассоциации.

Имеются многочисленные тесты, использование которых оправдано в конкретных ситуациях. Детальное рассмотрение и обоснование этих тестов является задачей математической статистики и выходит за рамки настоящей книги. Однако все тесты объединяет общий принцип: в основе расчетов лежит определенная функция, связывающая результирующий показатель с разницей между фактически наблюдаемыми значениями и теми значениями, которые можно было бы ожидать при справедливости нулевой гипотезы с учетом степени варибельности признака в выборке. Таким образом, любой тест, направленный на выявление статистической достоверности различий, можно рассматривать как дробь, числитель которой возрастает по мере возрастания разницы между фактическим и теоретически ожидаемым значением, основанным на нулевой гипотезе, а знаменатель отражает варибельность оценок в выборке и уменьшается по мере увеличения численности выборки. Все тесты приводят к суждению о вероятности P нулевой гипотезы. Значение P отражает вероятность того, что результаты, полученные в ходе исследования, связаны с исключительным влиянием случая, т. е. ассоциация между воздействием и изучаемым эффектом полностью отсутствует (нулевая гипотеза H_0 верна). Чем больше значение статистического теста, тем меньше величина P . Пороговое значение P , при котором результаты исследования можно считать достаточно достоверными, выбирается произвольно. В большинстве исследований принято говорить о достоверности результатов при пороговом значении $P = 0,05$. Таким образом, значение $P < 0,05$ позволяет утверждать, что при выявленном уровне ассоциации между изучаемым воздействием и эффектом или при еще более выраженной ассоциации вероятность случайности такой ассоциации не превышает 5 %. Напротив, при значении $P > 0,05$ нельзя исключить влияния случая как причины ассоциации, и нулевая гипотеза не может быть отвергнута, а следовательно, результаты не являются статистически значимыми (валидными).

Обычно величина P , полученная на основе статистического теста, отражает степень различий между двумя изучаемыми группами, независимо от направленности этих различий. В подобных случаях говорят о **двусторонних критериях** (*two-sided, two-tailed*), которые

чаще всего используются в исследованиях, так как позволяют оценить различия между выборками, независимо от направленности эффекта, связанного с изучаемым воздействием. Однако возможно использование и **односторонних критериев** (*one-sided, one-tailed*), когда предполагается наличие сведений о направленности связанного с воздействием эффекта в сопоставляемых группах. Использование односторонних критериев оправдано лишь тогда, когда имеется сильно аргументированная априорная гипотеза, отражающая предполагаемый эффект воздействия, а цель исследования — лишь получить дополнительные сведения о величине этого эффекта при его заведомо известной направленности, а также в тех случаях, когда исследование проводится с заранее заданной целью опровергнуть результаты предыдущего исследования.

В описании методики статистического анализа всегда следует представлять сведения о том, использовался в данном исследовании одно- или двусторонний критерий, что позволит более точно оценить результаты.

Несмотря на концептуальную схожесть статистических тестов, используемых для проверки гипотезы, не все они в равной степени приемлемы в конкретных ситуациях. Выбор того или иного теста во многом основан на особенностях анализируемой гипотезы и характере данных. В клинических исследованиях чаще всего используются *t*-критерий Стьюдента для анализа непрерывных количественных показателей, подчиняющихся нормальному распределению, и его аналог для показателей, не подчиняющихся нормальному распределению, — непараметрический критерий Манна—Уитни. При сопоставлении парных выборок часто используют парный *t*-критерий Стьюдента (если непрерывные показатели подчиняются нормальному распределению) или его аналог для показателей, не подчиняющихся нормальному распределению, — непараметрический критерий Уилкоксона. Для анализа дискретных и качественных показателей чаще всего используют критерий χ^2 или точный метод Фишера. Принципы расчетов *t*-критерия Стьюдента и критерия χ^2 рассматриваются ниже, детальную информацию об иных статистических тестах можно найти в специальных руководствах (Наследов А. Д., 2004; Сидоренко Е. В., 2007; Rothman K. J. [et al.], 2008; Peat J. K., Barton V., 2014; Woodward M., 2014). Большая часть прикладных статистических программ (SPSS, SAS, STATA, STATISTICA, EpiInfo) позволяет использовать названные критерии. В общем виде алгоритм выбора соответствующих статистических критериев представлен на рис. 48.

Для непрерывных значений показателей проверка нулевой гипотезы часто сводится к определению статистической достоверности выявленной разности между средними значениями в двух сопостав-

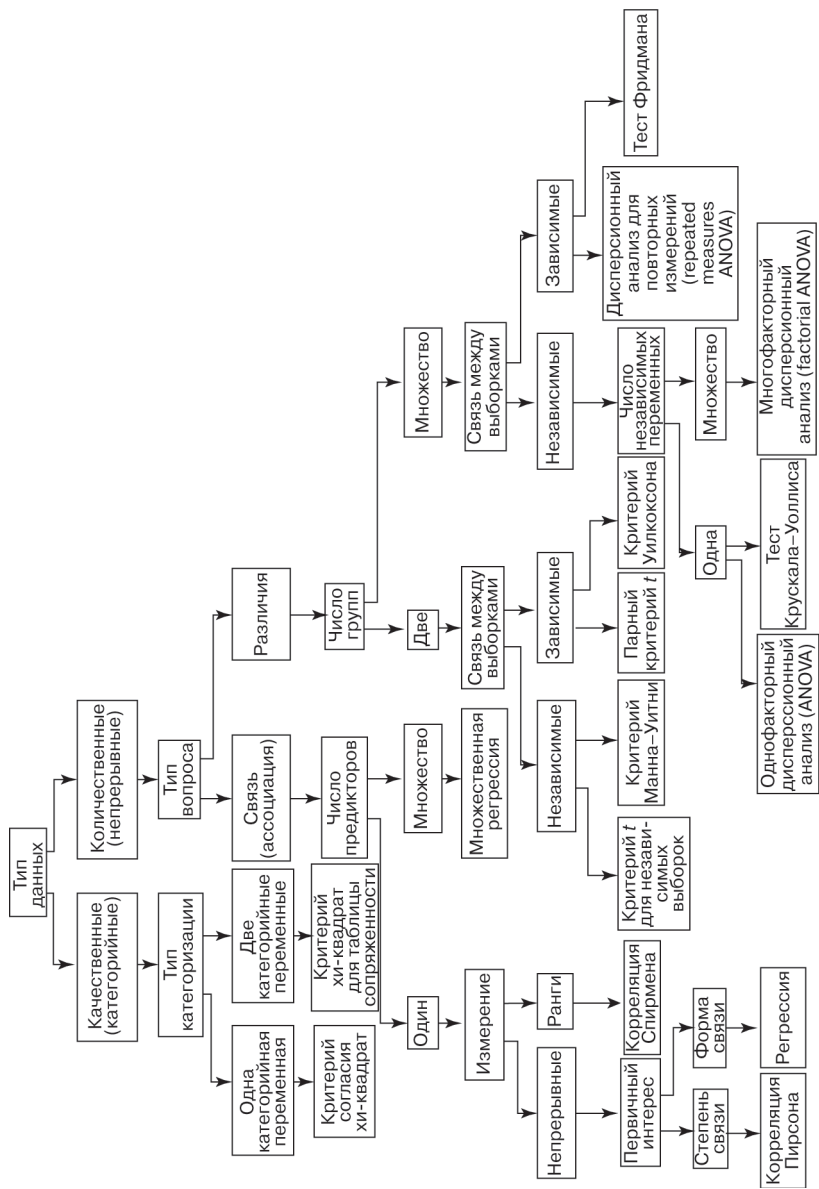


Рис. 48. Алгоритм выбора статистического критерия (Howell D. C., 2013)

ляемых группах. Если значения в сопоставляемых выборах подчиняются нормальному (гауссовскому) распределению, что, как правило, требует достаточно большой численности выборки, часто используют t -критерий Стьюдента. Если обозначить средние значения показателей в двух сопоставляемых группах как \bar{X}_1 и \bar{X}_0 , то нулевая гипотеза, предполагающая отсутствие различий между сопоставляемыми выборками, будет иметь следующее математическое выражение:

$$H_0: \bar{X}_1 = \bar{X}_0.$$

Альтернативная гипотеза имеет вид:

$$H_1: \bar{X}_1 \neq \bar{X}_0.$$

Для оценки достоверности различий вычисляют:

$$t = \frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{\left(\frac{S_p^2}{n_1} + \frac{S_p^2}{n_0}\right)}},$$

где t — значение t -распределения.

В данной формуле числитель отражает сопоставление разницы между средними значениями показателя в двух группах ($\bar{X}_1 - \bar{X}_0$) и значением этой разницы, которое можно было бы ожидать в случае истинности нулевой гипотезы, т. е. $\bar{X}_1 - \bar{X}_0 = 0$. В знаменателе отражены численность групп n_1 и n_0 , а также суммарная для обеих групп оценка вариации признака S_p^2 . Последняя вычисляется как взвешенная вариация признака, наблюдаемая в каждой группе (S_1^2 и S_0^2), с учетом численности наблюдений в каждой группе. Если обе группы имеют равную численность ($n_1 = n_0$), вес вариации признаков в каждой группе равен, а суммарная для обеих групп вариация является средней арифметической вариации в обеих группах:

$$S_p^2 = (S_1^2 + S_0^2)/2.$$

Если группы не являются равночисленными ($n_1 \neq n_0$), суммарная вариация вычисляется следующим образом:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_0 - 1)S_0^2}{(n_1 + n_0 - 2)}.$$

Вычисленное значение t следует сопоставить с критическим значением распределения t с числом степеней свободы: $df = n_1 + n_0 - 2$ (прилож. 2).

Пример

Рассмотрим данные о связи курения матери во время беременности с массой тела новорожденных детей (Kelmanson I. A., 2009). У 63 детей, мате-

ри которых курили во время беременности (n_1), среднее значение массы тела при рождении $\bar{X}_1 = 2800$ г, $\sigma = 698$ г. У 137 новорожденных детей, матери которых не курили во время беременности (n_0), среднее значение массы тела при рождении $\bar{X}_0 = 3150$ г, $\sigma = 773$ г.

Нулевая гипотеза H_0 может быть сформулирована следующим образом: отсутствуют различия массы тела у детей, родившихся у курящих и некурящих матерей ($\bar{X}_1 = \bar{X}_0$).

Для оценки вероятности нулевой гипотезы воспользуемся t -критерием Стьюдента. Расчеты выглядят следующим образом:

$$s_p^2 = [(63 - 1) \times 698^2 + (137 - 1) \times 773^2] / (63 + 137 - 2) = 562\,982,8.$$

$$t = (2800 - 3150) / \sqrt{[(562\,982,8/63) + (562\,982,8/137)]} = \\ = -350 / \sqrt{[8936,2 + 4109,4]} = 350 / 114,2 = 3,065.$$

Для определения значения P , соответствующей полученной величине t -критерия, следует воспользоваться таблицей t -распределения при числе степеней свободы, равном 198 (прилож. 2). Знак «минус» можно не учитывать (его бы и не было, если бы мы поменяли местами \bar{X}_1 и \bar{X}_0). Из таблицы следует, что даже для числа степеней свободы, равного 100, вычисленное значение t превышает критическое значение 2,626, соответствующее вероятности $P = 0,01$ для двустороннего критерия. Вероятность нулевой гипотезы (т. е. того, что выявленные различия между двумя группами связаны исключительно со случайностью) составляет менее 0,01. Следовательно, с вероятностью выше 99 % нулевая гипотеза может быть отвергнута, а значит, различия значений массы тела в двух сопоставляемых группах с вероятностью выше 99 % реально существуют. Такой уровень вероятности позволяет говорить о статистической достоверности полученных результатов.

В случае сопоставления дискретных переменных наиболее распространенным критерием значимости ассоциации является статистика χ^2 . В общем виде формула для расчета данного показателя выглядит следующим образом:

$$\chi^2(df) = \sum \frac{(O - E)^2}{E},$$

где O — фактическое число наблюдений в данной категории; E — теоретически ожидаемое число, рассчитанное исходя из нулевой гипотезы; \sum — знак суммирования значений для всех категорий; df — число степеней свободы, связанное с данной статистикой.

Ожидаемые для каждой категории значения определяются исходя из нулевой гипотезы и предположения, что пропорция выявления анализируемого фактора одинакова в сопоставляемых группах. Для

выполнения расчетов весьма удобна таблица вида 2×2 (табл. 21). Значение a — это число случаев распознаваемого патологического состояния среди лиц, подвергшихся патогенному воздействию. Если допустить отсутствие ассоциации между патогенным воздействием и риском развития заболевания, то следует ожидать, что доля лиц, страдающих заболеванием и подвергшихся предшествовавшему патогенному воздействию, будет равна общей доле лиц в популяции, подвергшихся патогенному воздействию. В данном случае теоретически ожидаемое число лиц, имеющих заболевание и подвергшихся предшествовавшему патогенному воздействию, будет вычисляться как произведение доли лиц в популяции, подвергшихся воздействию: $(a + b)/T$, и числа лиц, имеющих признаки заболевания: $(a + c)$:

$$\frac{(a + b)(a + c)}{T}.$$

Таблица 21

Представление данных в четырехпольной таблице

Воздействие	Наличие эффекта заболевания		Суммарно
	Да	Нет	
Имелось воздействие	a	b	$a + b$
Отсутствовало воздействие	c	d	$c + d$
Суммарно	$a + c$	$b + d$	$T = a + b + c + d$

По сходному принципу будут вычисляться и ожидаемые значения в остальных клетках четырехпольной (тетрафорической) таблицы (табл. 22).

Таблица 22

Ожидаемые значения в четырехпольной таблице, рассчитанные исходя из предположения случайности ассоциации

Воздействие	Наличие эффекта заболевания	
	Да	Нет
Имелось воздействие	$[(a+b)(a+c)]/T$	$[(a + b)(b + d)]/T$
Отсутствовало воздействие	$[(c+d)(a+c)]/T$	$[(c + d)(b + d)]/T$

Для определения значения P на основе вычисленного показателя χ^2 необходимо сопоставление полученного значения со значениями распределения χ^2 в зависимости от числа степеней свободы (прилож. 3). Для таблицы вида: $r \times c$ число степеней свободы определяется как произведение: $(r - 1) \cdot (c - 1)$. Очевидно, для любой таблицы вида 2×2 , число степеней свободы составит $(2 - 1) \cdot (2 - 1) = 1$. Эта величина означает, что в таблице вида 2×2 при фиксированных мар-

гинальных значениях достаточно сведений лишь о показателе одного поля таблицы для восстановления значений в оставшихся трех полях. Значение P , соответствующее распределению χ^2 с одной степенью свободы, можно также вычислить исходя из того, что квадратный корень из χ^2 соответствует стандартному нормальному распределению с двусторонним критерием.

Пример

Использование критерия χ^2 для расчета статистической значимости различий может быть проиллюстрировано примером связи курения матери во время беременности с риском нарушений засыпания у детей первого года жизни (Kelmanson I. A., 2009) (табл. 23). В соответствии с нулевой гипотезой H_0 распределение частоты нарушений сна у детей, родившихся у матерей, куривших и не куривших во время беременности, не отличается (этой частоте соответствует ожидаемое распределение случаев). Расчеты будут выглядеть следующим образом:

$$\chi^2_{(df=1)} = (15 - 9)^2/9 + (48 - 54)^2/54 + (14 - 20)^2/20 + (123 - 117)^2/117 = 6,78.$$

Таблица 23

Связь курения матери во время беременности с нарушениями сна у детей первого года жизни (Kelmanson I. A., 2009)

Курение матери во время беременности	Случаи нарушений сна у детей		Суммарно
	Да	Нет	
<i>Фактически наблюдаемые</i>			
Курила	$a = 15$	$b = 48$	$a + b = 63$
Не курила	$c = 14$	$d = 123$	$c + d = 137$
Суммарно	$a + c = 29$	$b + d = 171$	$T = 200$
<i>Теоретически ожидаемые значения</i>			
Курила	$[(a + b)(a + c)]/T = 9$	$[(a + b)(b + d)]/T = 54$	63
Не курила	$[(c + d)(a + c)]/T = 20$	$[(c + d)(b + d)]/T = 117$	137
Суммарно	29	171	200

Данной величине χ^2 при числе степеней свободы, равном 1, соответствует вероятность нулевой гипотезы $P < 0,01$ (критическое значение для указанной вероятности равно 6,635) (прилож. 3). Это означает, что с вероятностью более 99 % нулевая гипотеза может быть отвергнута, а значит, имеет место статистическая достоверность выявленных различий.

Имеются и модификации формулы для расчета значения χ^2 :

$$\chi^2(df) = \frac{\left[a - \frac{(a+b) \cdot (a+c)}{T} \right]^2}{\frac{(a+b) \cdot (c+d)(a+c)(d+b)}{T^2(T-1)}},$$

а также:

$$\chi^2(df) = (ad - bc)^2 \frac{T}{(a+b)(c+d)(a+c)(b+d)}.$$

В ряде случаев число наблюдений, соответствующих ячейкам четырехпольной таблицы, относительно невелико. Если фактические или теоретически ожидаемые значения, соответствующие отдельным ячейкам таблицы, меньше 5, а также в тех случаях, когда суммарное значение наблюдений во всех ячейках T меньше 20, при расчете критерия χ^2 рекомендуется осуществлять поправку на непрерывность распределения, предложенную F. Yates. Она получила наименование «**поправка Йейтса**» (*Yates's correction*). С учетом указанной поправки формула для расчета χ^2 выглядит следующим образом:

$$\chi^2(df) = (ad - bc - 0,5T)^2 \frac{T}{(a+c)(b+d)(a+b)(c+d)}.$$

Ряд авторов в подобных случаях отдает предпочтение использованию точного метода Фишера, методика расчета которого излагается в специальных руководствах (Наследов А. Д., 2004; Сидоренко Е. В., 2007; Armitage P. [et al.], 2001; Peat J. K., Barton B., 2014; Woodward M., 2014).

4.2.2. *Оценочные методы: определение доверительных интервалов*

Как явствует из сути любого статистического теста, направленного на определение достоверности различий, величина показателя P является функцией двух параметров: выраженности различий между сопоставляемыми группами (или, что эквивалентно, выраженности силы ассоциации) и численности выборки. Следовательно, даже весьма незначительные различия между выборками могут оказаться статистически достоверными при условии их большой численности, и, наоборот, выраженные различия могут оказаться статистически недостоверными, если численность выборки мала. В этой связи ве-

личина P должна рассматриваться как руководство к действию, а не как жесткий критерий, на основании которого можно делать безусловный вывод о наличии или отсутствии изучаемых различий и ассоциаций. В исследованиях всегда целесообразно представлять полученные значения P , а не ограничиваться указанием на достоверность или недостоверность ассоциации. Так, значения P , равные 0,07 и 0,7, с формальной точки зрения будут означать отсутствие статистической достоверности различий, если в качестве порогового уровня принять значение $P = 0,05$. Однако значение $P = 0,07$ значительно ближе к уровню статистической значимости, чем $P = 0,7$, и можно ожидать, что при увеличении численности выборки оно могло бы достигнуть уровня статистической достоверности. В результате этого, возможной практической рекомендацией может быть продолжение исследования на иной выборке адекватной численности.

Чтобы избежать некоторой трудности интерпретации, связанной с тем, что значение P зависит от силы ассоциации и числа наблюдений, можно прибегнуть к логически связанному со значением P , но более информативному показателю, отражающему влияние случая на результаты исследования. Таким показателем является **доверительный интервал (ДИ)** (*confidence interval, CI*). ДИ представляет собой диапазон, в пределах которого с заданной вероятностью находится истинное значение исследуемого эффекта.

Пример

Курение матери во время беременности ассоциировано с более низкими значениями массы тела новорожденных, и среднее значение массы тела у детей, родившихся у курящих женщин, на 350 г меньше, чем у детей, родившихся у некурящих матерей (Kelmanson I. A., 2009). Однако более детальную информацию дают значения 95 %-го ДИ этой разницы, составившей от 574 до 126 г. Масса тела новорожденных детей, родившихся у женщин, куривших во время беременности, ниже, чем у детей, родившихся у некурящих женщин, причем с вероятностью 95 % можно говорить о том, что истинное значение разности средних показателей массы тела этих детей не менее 126 г и не более 574 г.

Сведения о ДИ могут компенсировать необходимость представления сведений о величине P при оценке значимости выявленных различий или ассоциаций с заданной вероятностью. Очевидно, если значение $P > 0,05$, то 95 %-й ДИ вокруг разницы средних показателей будет захватывать значение, равное нулю, что по смыслу соответствует отсутствию достоверных различий. Напротив, если 95 %-й ДИ не захватывает значения, равного нулю, то можно говорить о достоверных различиях показателей с $P < 0,05$. В примере можно сделать вывод о том, что выявленная ассоциация достоверна с $P < 0,05$, так как минимальное значение 95 %-й ДИ вокруг разницы средних показателей выше нуля.

Кроме того, определение ДИ позволяет оценить вариабельность вычисленных показателей ассоциации, непосредственно связанных с численностью выборки: чем больше численность выборки, тем при прочих равных условиях более узкими будут границы ДИ. Сведения о диапазоне ДИ особо интересны в случае выявления недостоверных результатов исследования ассоциации: узкие границы ДИ будут скорее свидетельствовать об истинном отсутствии ассоциации, тогда как широкий диапазон может свидетельствовать о том, что для констатации эффекта, вероятно, необходима большая численность выборки, что позволит увеличить статистическую силу исследования и исключить возможное влияние случайности на результаты.

Таким образом, значение P и границы ДИ дают максимальную информацию о возможном влиянии шанса на результаты исследования.

Формула для расчета ДИ вокруг показателей в общем виде выглядит следующим образом:

$$ДИ = d \pm z \cdot СОС,$$

где d — абсолютное значение показателя; z — соответствующее заданному уровню достоверности значение плотности нормального распределения (табл. 24).

Для расчета 95 %-го ДИ с соответствующим ему значением $P = 0,05$ величина z составляет 1,96. Расчеты 95 %-го ДИ встречаются чаще всего, что, однако, не исключает возможности определения границ с иным уровнем статистической значимости. Если требуется расчет 99 %-го ДИ с соответствующим ему значением $P = 0,01$, то $z = 2,5758$. Если же требуется расчет 99,9 %-го ДИ с соответствующим ему значением $P = 0,001$, то $z = 3,2905$. Доверительные границы при таких расчетах будут более широкими, чем для 95 %-го ДИ.

Таблица 24

Критические (двусторонние) значения стандартного нормального распределения. Представлены значения z , соответствующие проценту P стандартного нормального распределения, находящемуся за пределами диапазона от $-z$ до $+z$ (Woodward M., 2014)

P	z	P	z
90	0,1257	10	1,6449
80	0,2533	5	1,9600
70	0,3853	2	2,3263
60	0,5244	1	2,5758
50	0,6745	0,2	3,0902
40	0,8416	0,1	3,2905
30	1,0364	0,02	3,7190
20	1,2816	0,01	3,8906
15	1,4395		

Очевидно, различия в расчетах будут связаны с методами вычисления СОС. В тех случаях, когда изучаемый показатель представляет собой частоту конкретного события и объектом изучения является одна группа (например, при исследовании популяционных срезов с изучением показателей смертности, заболеваемости и т. п.), СОС определяется следующим образом:

$$\text{СОС} = \sqrt{\frac{p(1-p)}{n}},$$

где p — частота; n — число наблюдений.

Пример

Требуется определить ДИ вокруг регистрируемой частоты нарушений сна у детей, родившихся у некурящих матерей (см. табл. 23). Нарушения сна были зафиксированы у 14 из 137 детей, родившихся у не куривших матерей ($n = 137$). Таким образом, $p = 14/137 = 0,102$ (или 10,2 %). Тогда:

$$\text{СОС} = \sqrt{[0,102(1 - 0,102)/137]} = 0,026.$$

$$95\% \text{-й ДИ} = 0,102 \pm 1,96 \times 0,026 = 0,051 \div 0,153.$$

Интерпретация полученных данных такова: истинное значение частоты, с которой нарушения сна выявляются у детей первого года жизни, родившихся у не куривших во время беременности матерей, с 95 % вероятностью находится в диапазоне от 5,1 до 15,3 %.

Если сопоставляется разность частоты события в двух группах, расчет СОС будет проводиться следующим образом:

$$\text{СОС} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_0(1-p_0)}{n_0}},$$

где p_1 и p_0 — частота события в 1-й и 2-й группах, соответственно, а n_1 и n_0 — число наблюдений в 1-й и 2-й группах соответственно.

Пример

Сопоставляя частоту выявления нарушений сна у детей первого года жизни, родившихся у курящих (группа 1) и некурящих (группа 2) матерей (см. табл. 23), можно отметить, что в группе 1 частота составила: $p_1 = 15/63 = 0,238$ (23,8 %); в группе 2 частота составила: $p_0 = 14/137 = 0,102$ (10,2 %). Разность частот ($p_1 - p_0$) составляет: $0,238 - 0,102 = 0,136$ (13,6 %). Оценим 95 %-й ДИ вокруг этой разности:

$$\begin{aligned} \text{СОС} &= \sqrt{[0,238(1 - 0,238)/63] + [0,102(1 - 0,102)/137]} = \\ &= \sqrt{(0,00288 + 0,000669)} = 0,0596; \end{aligned}$$

$$95\% \text{-й ДИ} = 0,136 \pm 1,96 \times 0,0596 = 0,0192 \div 0,253.$$

Следовательно, истинное значение разницы частоты нарушений сна в сопоставляемых группах с 95 % вероятностью находится в диапазоне от 1,9 до 25,3 %.

Заслуживает интереса подход к расчету 95 %-го ДИ для непрерывных количественных показателей. Расчеты ДИ для среднего значения, полученного в одной обследуемой группе, достаточно просты. СОС вычисляется следующим образом:

$$COС = \frac{CO}{\sqrt{n}},$$

где CO – среднеквадратическое отклонение; n – число наблюдений. Тогда 95%-й ДИ будет вычисляться следующим образом:

$$95\% \text{-й ДИ} = \bar{X} \pm 1,96 \times COС.$$

Для расчета 95 %-го ДИ вокруг разности средних значений в двух независимых выборках используются следующие формулы:

$$COС = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_0 - 1)S_0^2}{(n_1 + n_0 - 2)} \left(\frac{1}{n_1} + \frac{1}{n_0} \right)},$$

или

$$COС = \sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_0} \right)}.$$

Тогда 95 %-й ДИ разности средних значений \bar{X}_1 и \bar{X}_0 определяется:

$$95\% \text{-й ДИ} = (\bar{X}_1 - \bar{X}_0) \pm 1,96 \times COС.$$

В данной формуле S_1^2 и S_0^2 – вариация признака в группах 1 и 2, соответственно; S_p^2 – суммарная для двух групп вариация признака; $\bar{X}_1 - \bar{X}_0$ – разность средних значений признака в группах 1 и 2; n_1 и n_0 – число наблюдений в группах 1 и 2.

Весьма полезным может быть графическое представление точечного показателя (*point estimate*) и границ 95 %-го ДИ. Нагляднее всего это можно осуществить при помощи специальной диаграммы (в качестве примера взяты данные о разнице массы тела новорожденных, родившихся у куривших и не куривших во время беременности женщин) (рис. 49).

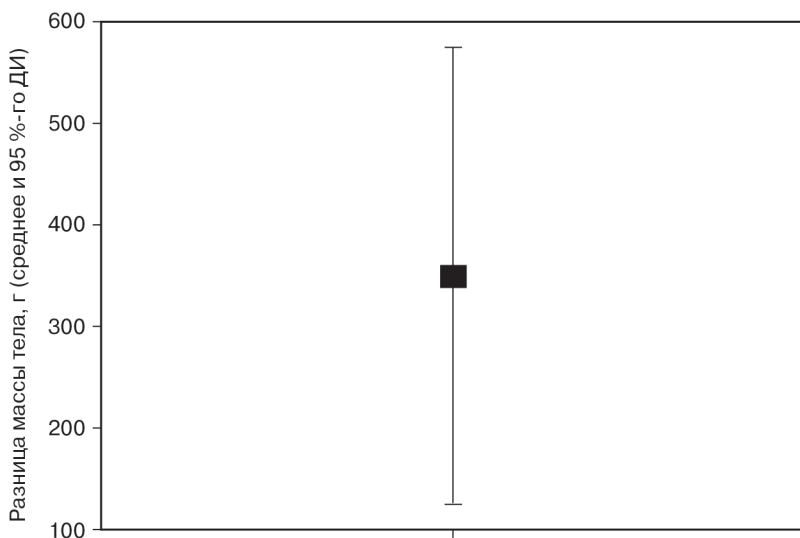


Рис. 49. Графическое представление разницы средних показателей и 95 %-го ДИ этой разницы

Интерпретируя результаты статистической проверки гипотезы и оценочных методов, следует иметь в виду ряд принципиальных положений:

- Нельзя рассматривать значение P как неопровержимый аргумент при решении вопроса о случайности выявленной ассоциации. Даже самое малое значение P еще не исключает возможной случайности ассоциации, а большая величина P не означает полного отсутствия реальной связи между изучаемыми явлениями. Во всех случаях речь может идти лишь о большей или меньшей вероятности того, что обнаруженные закономерности являются результатом простого совпадения. Вот почему всякий раз полученное абсолютное значение P следует анализировать в совокупности с результатами оценки ДИ и численности обследуемой группы.

- Следует помнить о том, что вопрос о статистической значимости ассоциации не должен подменять вопроса о клинико-биологической обоснованности выявленной ассоциации.

- Необходимо иметь в виду, что значения P и вычисляемые диапазоны ДИ позволяют судить лишь о достоверности обнаруженной ассоциации, но не дают информации о возможном влиянии на полученные результаты систематической ошибки и искажающих факторов.

В ходе исследований нередко собирают информацию, касающуюся многочисленных потенциальных факторов риска того или иного заболевания или патологического состояния. В подобных случаях часто выполняются множественные статистические тесты, направленные на выявление ассоциации между заболеванием и каждым из потенциальных факторов риска. Возможна и обратная ситуация, когда исследуются возможные множественные последствия одного предполагаемого воздействия, при которых также выполняются множественные статистические тесты. Во всех указанных случаях имеют место **множественные сравнения**. Если вероятность неверности одного статистического вывода равна α , то при построении m статистических выводов вероятность того, что хотя бы один из них будет неверным, существенно возрастает и составляет:

$$1 - (1 - \alpha)^m.$$

Иными словами, по мере увеличения числа параметров, оцениваемых при помощи статистических тестов в ходе множественных сравнений, возрастает и вероятность влияния случая на конечные результаты. Таким образом, если мы хотим, чтобы вероятность наличия хотя бы одного неверного вывода при проведении m сравнений была равна 0,05, то достаточно установить вероятность неверного отклонения нулевой гипотезы, равной $0,05/m$ для каждого вывода. Такой подход именуется **поправкой Бонферрони** (*Bonferroni correction*) (Woodward M., 2014). Согласно этой поправке, если, например, выполняется 5 сравнений, то за критический уровень P , позволяющий в каждом отдельном сравнении говорить о статистической значимости, следует принять не 0,05, а $0,05/5$, т. е. 0,01. Это позволит обеспечить общий уровень достоверности, равный 0,05, при 5 сравнениях.

Любая выявленная ассоциация, не предполагавшаяся на стадии планирования исследования, требует очень внимательного и осторожного отношения, и в идеальном случае нуждается в перепроверке в ходе последующих специально организованных исследований.

4.3. Выраженность эффекта

В ряде случаев в ходе проведения исследования бывает недостаточно просто констатировать наличие или отсутствие эффекта как такового; часто этот эффект требуется каким-то образом «измерить». При этом говорят о **выраженности эффекта** (*effect size*) (Cohen B. H., Lea R. B., 2004; Howell D. C., 2013; Kline R. B., 2013; Peat J. K., Barton B., 2014). Когда в качестве критериев выраженности эффекта используются непрерывные количественные показатели, первым этапом статистических расчетов будет решение вопроса

о том, какому показателю следует отдать предпочтение: прямому или так называемому стандартизованному.

Прямой показатель выраженности эффекта (*unstandardized effect size*) представляет собой непосредственное значение разницы средних величин (*mean difference*), наблюдавшихся в двух сопоставляемых группах.

Стандартизованный показатель выраженности эффекта (*standardized effect size*) представляет собой разницу средних значений, стандартизованную по показателю среднеквадратического отклонения. Поскольку среднеквадратическое отклонение имеет ту же размерность, что и среднее, полученное стандартизованное значение утрачивает размерность, которая была присуща исходному показателю разности средних значений. Выраженность эффекта в таком случае дана в долях среднеквадратического отклонения. Если предполагается, что исходные показатели подчинялись законам нормального распределения, то стандартизованные показатели подчиняются z -распределению и представляют собой шкалу z (*z-score*).

Показатель Коэна d (*Cohen's d*) рассчитывается как разность двух средних показателей, деленная на обобщенный показатель СО:

$$d = \frac{\bar{X}_1 - \bar{X}_0}{CO}$$

Чаще всего он используется при сопоставлении средних значений в двух выборках с непрерывными показателями в качестве меры стандартизованной выраженности эффекта. Если вариация признака в обеих группах одинакова, можно использовать значение СО любой из двух групп. Если же вариация признака различная, то рассчитывают обобщенный показатель СО для двух групп:

$$CO = \sqrt{\frac{(CO_1^2 + CO_0^2)}{2}}$$

Показатель d имеет асимптотически нормальное распределение, т.е. он тем ближе к нормальному распределению, чем больше численность обследуемой выборки. Вариация d оценивается следующим образом:

$$Var(d) = \frac{n_1 + n_0}{n_1 n_0} + \frac{d^2}{2(n_1 + n_0)}$$

Нижняя и верхняя границы 95 %-го ДИ показателя d будут определяться следующим образом:

$$95 \% - \text{й ДИ} = d \pm 1,96\sqrt{\text{Var}(d)}$$

В соответствии с принятыми рекомендациями, значение $d \leq 0,2$ интерпретируется как малая выраженность эффекта, $d = 0,3-0,5$ — как средняя выраженность эффекта, а $d \geq 0,8$ — как сильная выраженность эффекта (Barker C. [et al.], 2002; Cohen B. H., Lea R. B., 2004; Peat J. K., Barton V., 2014). При этом нужно понимать, что интерпретация показателя d должна рассматриваться в контексте проводимых исследований.

Преимущества

— Стандартизованный показатель выраженности эффекта не зависит от численности выборки.

— Исследование стандартизованных показателей позволяет сопоставлять выраженность эффектов, полученных при различных измерениях, даже если эти измерения проводились с использованием различных методик, шкал и даже если исследования проводились в различных выборках.

Пример

Предметом исследования является изучение выраженности симптомов эмоциональных расстройств и нарушений поведения у 40 детей дошкольного возраста, родившихся в исходе кесарева сечения (основная группа). Контрольную группу составили 40 детей того же возраста и пола, родившиеся в исходе естественных родов. Нарушения поведения оцениваются по опроснику Ахенбаха, заполняемому воспитателями детского сада (Kelmanson I. A., 2013b). Выявлено, что по шкале «Тревога/депрессия» дети основной группы имели в среднем $\bar{X}_1 = 2,28$ балла ($\bar{X}_1 = 1,09$), а дети из контрольной группы в среднем $\bar{X}_0 = 0,90$ балла ($\sigma_0 = 0,93$). По шкале «Нарушения сна» дети из основной группы имели в среднем $\bar{X}_1 = 2,48$ балла ($\sigma_1 = 1,52$), а в контрольной группе в среднем $\bar{X}_1 = 0,78$ балла ($\sigma_1 = 0,92$). По показателям обеих шкал между детьми двух групп имелись статистически достоверные различия ($P < 0,001$). Однако насколько велики выявленные эффекты и по какой шкале выраженность эффекта была большей?

Выраженность эффекта рассчитывается:

— для шкалы «Тревога/депрессия»:

$$d = (2,28 - 0,90) / \sqrt{[(1,09^2 + 0,93^2) / 2]} = 1,4;$$

— для шкалы «Нарушения сна»:

$$d = (2,48 - 0,78) / \sqrt{[(1,52^2 + 0,92^2) / 2]} = 1,35.$$

Можно сделать вывод, что эффекты по обеим шкалам были сильно выраженными и практически одинаковыми.

Таким образом, прямые (не стандартизованные) оценки выраженности эффекта можно использовать лишь в тех случаях, когда не вызы-

вает никаких сомнений полная сопоставимость показателей, полученных в отдельных исследованиях. В остальных случаях целесообразно прибегать к использованию стандартизованных показателей.

Показатель Коэна может давать завышенные оценки выраженности эффекта, если численность двух групп невелика. Поэтому при малой численности выборки рекомендуется рассчитывать **показатель Хэджеса g** (*Hedges's g*) (Cohen В. Н., Lea R. В., 2004; Peat J. К., Barton В., 2014):

$$g = \frac{(\bar{X}_1 - \bar{X}_0)}{\sqrt{\frac{(n_1 - 1)CO_1^2 + (n_0 - 1)CO_0^2}{(n_1 + n_0 - 2)}}}$$

Показатель g может быть рассчитан и на основе показателя d :

$$g = \frac{d}{\sqrt{\frac{(n_1 + n_0)}{(n_1 + n_0 - 2)}}}$$

Пример

Рассчитывается выраженность эффекта курения матери во время беременности в отношении массы тела новорожденного ребенка с учетом приведенных ранее данных: у 63 детей, матери которых курили во время беременности (n_1), среднее значение массы тела при рождении $\bar{X}_1 = 2800$ г, $CO_1 = 698$ г; у 137 новорожденных детей, матери которых не курили во время беременности (n_0), среднее значение массы тела при рождении $\bar{X}_0 = 3150$ г, $CO_0 = 773$ г.

Расчет выраженности эффекта с использованием показателя d дает следующие результаты:

$$\begin{aligned} \sigma &= \sqrt{(698^2 + 773^2)/2} = 736,5; \\ d &= (2800 - 3150)/736,5 = 0,48 \text{ (знак «минус» можно не учитывать);} \\ Var(d) &= [(63 + 137)/63 \cdot 137] + (0,48)^2/2 \cdot (63 + 137) = 0,023776; \\ \sqrt{Var(d)} &= 0,154; \\ 95\text{-й ДИ} &= 0,48 \pm 1,96 \cdot 0,154 = 0,18 \times 0,78; \\ g &= 0,48/\sqrt{[(63 + 137)/(63 + 137 - 2)]} = 0,478 \approx 0,48. \end{aligned}$$

Можно говорить о средней выраженности эффекта курения матери во время беременности в отношении снижения массы тела ребенка при рождении. При этом 95%-й ДИ показателя свидетельствует о том, что с 95% вероятностью эффект может находиться в диапазоне от малого (0,18) до среднего, приближающегося к сильному (0,78).

Интересно графическое представление данных о выраженности эффекта для большей наглядности. С этой целью чаще всего используется специальная диаграмма (*forest plot*), на которой представлены значения выраженности эффекта и границы 95%-го ДИ (рис. 50).

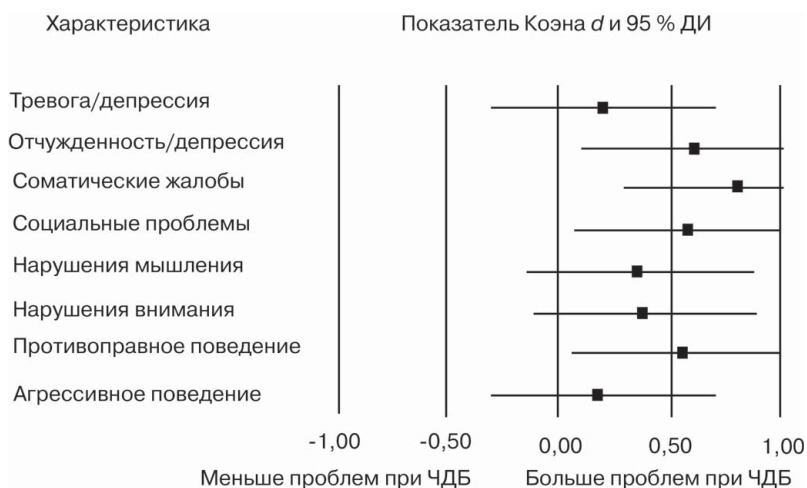


Рис. 50. Диаграмма, отражающая выраженность эффекта при сопоставлении особенностей поведения часто и длительно болеющих детей и здоровых детей из контрольной группы наблюдений (Kelmanson I. A., 2015)

Диаграмма взята из исследования, посвященного изучению особенностей поведения младших школьников, страдающих частыми и длительными респираторными заболеваниями (Kelmanson I. A., 2015). Статистически достоверные различия между часто и длительно болеющими детьми и детьми из контрольной группы наблюдений были выявлены по таким характеристикам поведения, как отчужденность/депрессия, наличие соматических жалоб, социальные проблемы и противоправное поведения. По всем перечисленным характеристикам показатель выраженности эффекта (d Коэна) был больше нуля, причем минимальные границы 95 %-го ДИ выраженности эффекта также превышали критический уровень, равный нулю. Это значит, что с вероятностью 95 % ($P < 0,05$) данные показатели различались в двух обследованных группах. Анализ диаграммы позволяет говорить о том, что по всем перечисленным показателям точечные значения выраженности эффекта соответствовали сильному эффекту и максимальная выраженность эффекта наблюдалась по показателю соматических жалоб. Вместе с тем границы 95%-го ДИ выраженности эффектов по этим показателям накладывались друг на друга, следовательно, нет убедительных оснований говорить о том, что какой-то один из выявленных эффектов существенно отличался от остальных. Границы 95 %-го ДИ достаточно широки, что позволяет предполо-

жить малую численность выборки. Действительно, было обследовано лишь 30 часто и длительно болеющих детей и 30 детей из контрольной группы. Таким образом, увеличение численности выборки могло бы сузить границы ДИ и позволило бы сделать более убедительное заключение о том, где максимально проявляется эффект.

Дисперсионный (вариационный) анализ (ANOVA) используется для изучения эффекта воздействия во многих исследованиях. Математическая идея ANOVA основана на соотношении межгрупповой и внутригрупповой частей дисперсии (изменчивости) изучаемой зависимой переменной. Внутригрупповая изменчивость рассматривается как обусловленная случайными причинами, а межгрупповая — как обусловленная действием изучаемого фактора на зависимую переменную. В общей изменчивости (дисперсии) зависимой переменной выделяются внутригрупповая (случайная) и межгрупповая (факторная) компоненты. Чем больше отношение межгрупповой изменчивости к внутригрупповой, тем больше различаются средние значения, соответствующие разным градациям фактора (Наследов А. Д., 2004). Нулевая гипотеза содержит утверждение о равенстве межгрупповой и внутригрупповой составляющих изменчивости и подразумевает альтернативное предположение о том, что межгрупповая составляющая изменчивости превышает внутригрупповую изменчивость. Нулевой гипотезе соответствует равенство средних значений зависимой переменной на всех уровнях фактора. Принятие альтернативной гипотезы означает, что по крайней мере два средних значения различаются (без уточнения, какие именно градации фактора различаются). Основные допущения дисперсионного анализа состоят в том, что распределения зависимой переменной для каждой градации фактора соответствуют нормальному закону, а дисперсии выборок, соответствующих разным градациям фактора, равны между собой. Кроме того, выборки, соответствующие градациям фактора, должны быть независимы (для межгруппового фактора) (Наследов А. Д., 2004).

Интересна оценка выраженности эффекта при использовании дисперсионного анализа. Чаще всего с указанной целью используют долю вариации признака η^2 (эта квадрат), связанную с влиянием фактора в общей вариации признака. Значения η^2 находятся в диапазоне от 0 до 1. Расчет этого показателя выглядит следующим образом:

$$\eta^2 = \frac{SS_B}{SS_T},$$

где SS_B — вариация признака, связанная с влиянием фактора; SS_T — общая вариация признака.

Показатель η^2 может быть преобразован в показатель Коэна f (Cohen's f), отражающий стандартизованную разницу средних значений в группах:

$$f = \frac{\sqrt{\eta^2}}{(1 - \eta^2)},$$

Показатель Коэна f выраженности эффекта: $f = 0,1$ (малая степень), $f = 0,25$ (средняя), $f = 0,4$ (сильная) (Peat J. K., Barton B., 2014).

При изучении **корреляции Пирсона** выраженность эффекта оценивается следующим образом: $r = 0,1$ оценивается как малый, $0,3$ — как средний и $0,5$ — как сильный эффект (Peat J. K., Barton B., 2014).

Показатель вариации коэффициента корреляции вычисляется следующим образом:

$$Var_r = \frac{(1 - r^2)^2}{n - 1},$$

где n — число наблюдений (Borenstein M. [et al.], 2009).

В тех ситуациях, когда осуществляется **регрессионный анализ**, обобщающим показателем является коэффициент регрессии R , эквивалентный по своей сути коэффициенту корреляции r . Квадрат коэффициента регрессии R^2 именуется **коэффициентом детерминации**; он отражает долю вариации зависимой переменной, которая может быть объяснена независимой, объясняющей переменной. Отчасти его можно рассматривать как меру выраженности эффекта, однако еще более информативным является показатель Коэна f , который может применяться как в случае простого, так и множественного регрессионного анализа. Расчет этого показателя в данном случае проводится так:

$$f = \sqrt{\frac{R^2}{(1 - R^2)}}.$$

Интерпретация этого показателя аналогична интерпретации показателя Коэна d и приведена выше.

Оценка выраженности эффекта в случаях качественных оценок. Как указывалось, результаты, основанные на применении бинарных критериев, лучше всего могут быть представлены в виде таблицы 2×2 . Строки будут отражать группу, в которой имело место изучаемое воздействие (основную), и контрольную группу, в которой указанное воздействие отсутствовало. Два столбца таблицы соответствуют двум вариантам исходов: наличию или отсутствию рассматри-

ваемого эффекта (см. табл. 21). Оценка эффекта изучаемого воздействия может быть основана на показателе **риска**.

Риск отражает вероятность наступления события, рассматриваемого как наличие эффекта, и рассчитывается как доля обследованных пациентов, у которых зафиксировано это событие. Риск наличия эффекта составит:

— в группе, подвергнутой воздействию:

$$r_E = a/(a + b);$$

— в контрольной группе:

$$r_C = c/(c + d).$$

Относительный риск (ОР) (*relative risk, RR*) будет определяться отношением риска наличия эффекта в группе, подвергнутой изучаемому воздействию, к аналогичному показателю в контрольной группе:

$$RR = r_E/r_C,$$

или

$$RR = \frac{\frac{a}{(a + b)}}{\frac{c}{(c + d)}} = \frac{a(c + d)}{c(a + b)}.$$

Относительный риск будет отражать эффект изучаемого воздействия по сравнению с **базовым риском** — риском возникновения эффекта в контрольной группе.

Расчет ДИ для показателей относительного риска. Следует иметь в виду, что распределение значений относительного риска отличается от нормального, оно характеризуется выраженной положительной асимметрией, так как значение относительного риска не может быть меньше нуля, но теоретически способно принимать любые положительные значения. Однако возможно преобразование такого распределения путем определения натурального логарифма (\ln) значений относительного риска, и полученные в результате трансформации значения будут характеризоваться распределением, близким к нормальному распределению. Такой прием, используемый в статистике, именуется лог-нормальным преобразованием. Следовательно, рассчитываются границы ДИ не вокруг непосредственно вычисленного значения относительного риска, а вокруг показателя, полученного в результате его лог-нормального преобразования. Чтобы вернуться к истинным значениям относительного риска, необходим расчет антилогарифмов вычисленных границ ДИ:

$$[\exp(\ln RR)], \text{ т. е. } e^{\ln RR}$$

Таким образом, 95 %-й ДИ для показателя относительного риска может быть рассчитан следующим образом:

$$\exp^{\ln RR \pm 1,96 \cdot COC},$$

где
$$COC = \sqrt{Var(\ln RR)}.$$

Пользуясь принятыми обозначениями (см. табл. 21) для расчета COC логарифма относительного риска ($\ln RR$), следует пользоваться формулой (Woodward M., 2014):

$$COC = \sqrt{\left(\frac{1}{a} + \frac{1}{c} - \frac{1}{(a+b)} - \frac{1}{(c+d)} \right)}.$$

Показатель относительного риска предполагает сопоставление частот возникновения тех или иных состояний в основной и контрольной группах. Следовательно, этот показатель можно применять в когортных обсервационных и экспериментальных исследованиях, где группы исходно формируются в зависимости от изучаемого воздействия, а в дальнейшем возможен учет частоты возникновения того или иного эффекта, связанного с этим воздействием.

Расчет показателя шансов. В исследованиях «случай – контроль», когда субъекты, входящие в исследование, подбираются в зависимости от наличия или отсутствия эффекта (заболевания, патологического состояния), как правило, не представляется возможным точно вычислить частоту развития этого заболевания в зависимости от наличия или отсутствия воздействия. Следовательно, формула, используемая для расчета показателя относительного риска, не подходит для вычислений в исследованиях типа «случай – контроль». Выраженность эффекта в такой ситуации можно оценить по показателю шансов. В отличие от показателя риска, шансы будут рассчитываться как отношение вероятности наступления события к вероятности его отсутствия. Так, шансы наступления эффекта составят:

– в группе, подвергнутой воздействию (основной):

$$r_E / (1 - r_E),$$

– в контрольной группе:

$$r_C / (1 - r_C),$$

Показатель отношения шансов (ОШ) (odds ratio, OR) является аналогом показателя относительного риска и отражает влияние изучаемого воздействия на шансы возникновения рассматриваемого эффекта по сравнению с базовыми шансами этого эффекта.

$$OR = [r_E / (1 - r_E)] / [r_C / (1 - r_C)],$$

Показатель отношения шансов может быть рассчитан как:

$$OR = \frac{ad}{cb}.$$

Проводится и расчет ДИ показателя отношения шансов. В данном случае также осуществляется лог-нормальное преобразование, и 95%-й ДИ для показателя отношения шансов может быть рассчитан следующим образом:

$$\exp^{\ln OR \pm 1.96 \cdot COC}$$

Формула для расчета COC логарифма отношения шансов выглядит так (Woodward M., 2014):

$$COC = \sqrt{\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}\right)}.$$

Показатель относительного риска связан с показателем отношения шансов. Если показатель базового риска (в контрольной группе) составляет r_C , то связь будет выглядеть следующим образом:

$$OR = \frac{RR(1 - r_C)}{1 - RR \cdot r_C}.$$

Значение показателя отношения шансов приближается к значению показателя относительного риска в тех случаях, когда базовый риск, т. е. частота изучаемого патологического состояния в популяции, невелик (менее 0,2) (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987). Действительно, если значение r_C стремится к нулю, то в приведенной выше формуле значение OR приближается к значению RR.

Показатель относительного риска лучше известен исследователям и проще для восприятия. Показатель отношения шансов порой труднее для интерпретации, хотя и основан на более строгих подходах к расчетам.

Если изучается показатель выраженности ассоциации между воздействием и эффектом без указания на то, какой конкретный показатель должен быть использован, предпочтение следует отдавать показателю отношения шансов, так как этот показатель более устойчив к влиянию численности выборки (Кельмансон И. А., 2002; Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008). В случае использования показателя относительного риска целесообразно убе-

даться в том, что вычисленный показатель относительного риска существенно не отличается от показателя отношения шансов.

Пример

В качестве примера расчета отношения шансов рассмотрим данные о связи курения матери во время беременности с риском нарушений сна ребенка первого года жизни (см. табл. 23).

Показатель составит:

$$OR = ad/cb = 15 \times 123 / 14 \times 48 = 2,7.$$

У детей, чьи матери курили во время беременности, риск развития нарушений сна в 2,7 выше, чем у детей, родившихся у некурящих женщин.

Рассчитаем границы 95 %-го ДИ показателя отношения шансов:

$$\ln OR = \ln 2,7 = 0,993;$$

$$СОС (\ln OR) = \sqrt{(1/15 + 1/48 + 1/14 + 1/123)} = \sqrt{0,167} = 0,409;$$

$$95 \% \text{-й ДИ } (\ln OR) = 0,993 \pm 1,96 \cdot 0,409 = 0,191 \div 1,795;$$

$$95 \% \text{-й ДИ } OR = 1,21 \div 6,02.$$

Риск нарушения сна у ребенка, ассоциированный с курением матери во время беременности, с 95 %-й вероятностью находится в диапазоне от 1,21 до 6,02.

Показатели относительного риска и отношения шансов могут иметь только положительные значения. Значение, равное 1, свидетельствует об отсутствии эффекта: риск на фоне воздействия, не увеличивается и не уменьшается. Значения, превышающие 1, показывают повышение вероятности развития эффекта. Значения ниже 1 свидетельствуют о снижении риска, и в подобной ситуации принято говорить о **протекторном отношении шансов** (*protective odds ratio*) (Peat J. K., Barton V., 2014). Подобная ситуация возможна, например, при оценке риска развития осложнений заболевания на фоне лечения. Признаком достоверности ассоциации является то, что 95%-й ДИ показателя относительного риска или показателя отношения шансов находится «по одну сторону» от критического значения, равного 1 (рис. 51). Так, в рассмотренном выше примере нижняя граница 95%-го ДИ показателя отношения шансов составила 1,21. Это значит, что с 95%-й вероятностью ($P < 0,05$) риск нарушений сна у ребенка увеличивается, если женщина курила во время беременности, так как даже наименьший показатель отношения шансов превышает 1. Имеет место высокая выраженность эффекта, оцениваемого по показателю отношения шансов, когда этот показатель равен 2 и больше, т. е. когда шансы появления эффекта на фоне рассматриваемого воздействия повышаются в два раза и более (Peat J. K., Barton V., 2014). Сходным образом, выраженный эффект очевиден, когда значение показателя отношения шансов составляет 0,5 и меньше, т. е. когда шансы появления эффекта на фоне рассматриваемого воздействия снижаются в два раза и более.



Рис. 51. Графическое представление показателя отношения шансов

Разница риска (PP) (*risk difference, RD*) – показатель, который может быть использован для оценки выраженности эффекта в исследованиях с бинарным откликом. Этот показатель относится к категории абсолютных, а не относительных показателей, демонстрирует, насколько риск развития эффекта в основной группе выше базового риска. Он рассчитывается следующим образом:

$$RD = r_E - r_C = [a/(a + b) - c/(c + d)].$$

Его стандартная ошибка:

$$COE = \sqrt{\frac{ab}{(a + b)^3} + \frac{cd}{(c + d)^3}}.$$

Показатель разницы риска имеет отрицательное значение при снижении вероятности наступления эффекта на фоне изучаемого воздействия, и в этих случаях говорят о снижении абсолютного риска. Показатель имеет положительное значение при повышении вероятности наступления эффекта на фоне воздействия. При отсутствии эффекта показатель разницы риска равен нулю. Нередко данный показатель именуют **атрибутивным риском (AR)** (*attributable risk, AR*), так как он отражает ту избыточность риска заболевания, которая непосредственно связана с рассматриваемым воздействием. При этом предполагается, что между рассматриваемым воздействием и эффектом существует реальная причинно-следственная связь. Если это условие выполняется, то показатель атрибутивного риска, представленный в процентах, будет отражать ту часть случаев забо-

левания среди лиц, подвергшихся рассматриваемому воздействию, которые можно было бы предотвратить за счет устранения этого патогенного воздействия. Эта характеристика именуется также **атрибутивной пропорцией, этиологической фракцией**:

$$AR (\%) = \left(\frac{AR}{r_c} \right) \cdot 100 \% = \frac{(r_e - r_c)}{r_e} \cdot 100 \%,$$

Если в когортных исследованиях можно получить сведения о появлении новых случаев (инциденте) заболевания в группе, подвергшейся изучаемому воздействию, то в большинстве исследований «случай – контроль» нет точных сведений об истинном инциденте заболевания среди лиц, подвергшихся и не подвергшихся рассматриваемому воздействию. Следовательно, нет возможности рассчитать показатель риска и вычислить атрибутивный риск. Однако в этих случаях процент атрибутивного риска можно рассчитать по модифицированной формуле, используя показатель отношения шансов (Hennekens C. H. [et al.], 1987; Rothman K. J. [et al.], 2008):

$$AR (\%) = \frac{(OR - 1)}{OR} \cdot 100 \%$$

В примере, отражающем связь между курением матери во время беременности и нарушениями сна у ребенка первого года жизни (см. табл. 23), относительный риск нарушений сна у ребенка, ассоциированный с курением матери, составил 2,7. Процент атрибутивного риска составит в данном случае:

$$AR (\%) = [(2,7 - 1)/2,7] \times 100 \% = 63 \%$$

Следовательно, если курение матери во время беременности является причинно значимым фактором нарушений сна у ребенка первого года жизни, то у 63 % грудных детей, имеющих нарушения сна и во внутриутробном периоде подвергшихся воздействию курения матери, нарушение сна связано с изолированным влиянием данного патогенного воздействия. Иными словами, у 63 % грудных детей, имеющих нарушения сна и подвергшихся влиянию курения матери во время беременности, данные нарушения сна можно было бы потенциально устранить, если бы женщины отказались от курения во время беременности.

Если изучаемое воздействие носит не патогенный, а позитивный, профилактический, саногенный характер, то разговор об атрибутивном риске теряет смысл. Однако представляется возможным вычислить по сути сходный показатель, отражающий позитивный эффект рассматриваемого воздействия, и в таком случае этот показатель именуется **превентивной фракцией** (*preventive fraction, PF*).

Очень интересным показателем, позволяющим улучшить клиническую интерпретацию выраженности эффекта, связанного с лечением, является такая характеристика, как **число пациентов, нуждающихся в лечении (ЧНЛ) для предотвращения хотя бы одного нежелательного проявления заболевания** (*number needed to be treated, NNT*). Данный показатель рассчитывается на основе показателя абсолютного снижения риска:

$$\text{ЧНЛ} = 1/\text{PP}.$$

Оценка выраженности эффекта позволяет сопоставить степень ассоциации между различными изучаемыми переменными, причем как в рамках одного исследования, так и при изучении результатов различных исследований. Однако возможны ситуации, когда ассоциации между различными изучаемыми переменными исследуются по-разному: в одном случае речь может идти о непрерывных переменных и средних значениях, в другом — о коэффициенте корреляции, в третьем — о бинарных откликах и показателях риска. Как сопоставить выраженности эффектов в таких случаях? Возможна взаимная «конвертация» показателей выраженности эффекта. Показатель отношения шансов *OR* может быть преобразован в показатель стандартизованной разности средних (*d* Коэна) следующим образом:

$$d = \ln \text{OR} \cdot \frac{\sqrt{3}}{\pi}$$

где π — математическая константа ($\pi \approx 3,14159$) (Borenstein M. [et al.], 2009). Вариация d (Var_d) будет равна:

$$\text{Var}_d = \text{Var}_{\ln \text{OR}} \cdot \frac{3}{\pi^2},$$

где $\text{Var}_{\ln \text{OR}}$ — вариация логарифма отношения шансов.

Возможно и преобразование коэффициента корреляции r в эквивалентное ему значение d Коэна (Borenstein M. [et al.], 2009):

$$d = \frac{2r}{\sqrt{1-r^2}}.$$

Вариация показателя d (Var_d), полученного из показателя r , будет рассчитываться следующим образом:

$$\text{Var}_d = \frac{4 \cdot \text{Var}_r}{(1-r^2)^3}.$$

4.4. Численность выборки и статистическая сила исследования

Численность выборки влияет на значение P и вероятность того, что любые выявленные различия достигнут уровня статистической значимости. Следовательно, на стадии планирования исследования возможна постановка вопроса о том, какова оптимальная численность обследуемой выборки (*sample size*), необходимая для констатации различий с заданным уровнем статистической значимости, если эти различия действительно имеют место.

Возможна и ситуация, в которой исследование проводится среди ограниченного, доступного круга лиц, и тогда после исследования требуется ответить на вопрос: насколько в обследованной выборке была вероятна констатация наличия различий с заданным уровнем значимости, если эти различия действительно существуют, т. е. речь идет об определении статистической силы исследования. При этом оценка будет проводиться задним числом (*post hoc*).

В отличие от оценки статистической достоверности, когда допускается нулевая гипотеза и задача исследователя состоит в попытке ее опровержения, при расчете численности выборки и статистической силы исследования нулевая гипотеза исходно отвергается, а истинной считается именно альтернативная гипотеза о существовании ассоциации между изучаемым воздействием и эффектом. Необходимо определить:

— Какова предположительная выраженность эффекта или каков тот минимальный, критический уровень выраженности эффекта, фиксация которого удовлетворит автора проводимого исследования?

— Направлено ли исследование на выявление даже малого эффекта или исследователь вполне удовлетворится констатацией лишь сильного эффекта?

Пример

При изучении связи курения матери во время беременности и нарушения сна у ребенка первого года жизни исследователи могут рассматривать значение отношения шансов 1,5 и более в качестве аргумента в пользу существования ассоциации. Этот уровень отражал бы увеличение вероятности расстройств сна у ребенка минимум на 50 %, если его мать курила во время беременности. В подобном случае нулевая гипотеза и альтернативная ей гипотеза об истинности ассоциации будут иметь математическое выражение:

$$\begin{aligned}H_0: OR &= 1,0; \\H_1: OR &> 1,5.\end{aligned}$$

На основе данных, полученных в ходе исследования, можно принять решение о возможности отвергнуть нулевую гипотезу в пользу

альтернативной. При этом возможны два корректных заключения:

— отвергается нулевая гипотеза при справедливости альтернативной гипотезы;

— невозможно отвергнуть нулевую гипотезу при ее истинности.

Однако возможны две ошибки в оценке истинности нулевой гипотезы:

— *Ошибка первого рода* возникает тогда, когда нулевая гипотеза отвергается, хотя на самом деле она верна. Вероятность ошибки первого рода равна величине P . Такая ошибка часто обозначается греческой буквой α (альфа). Так, при $\alpha = 0,05$ вероятность необоснованно отвергнуть нулевую гипотезу составляет 5 %, а $P = 0,05$.

— *Ошибка второго рода* возникает тогда, когда нулевая гипотеза необоснованно не отвергается, хотя на самом деле верна альтернативная гипотеза, утверждающая наличие истинных различий между исследуемыми группами. Вероятность ошибки второго рода обозначается греческой буквой β (бета) (табл. 25) (Кельмансон И. А., 2002; Власов В. В., 2004; Armitage P. [et al.], 2001; Woodward M., 2014).

Таблица 25

Возможные исходы, связанные с проверкой гипотезы

Заключение, основанное на результатах статистического теста	Истинная ситуация	
	H_0 верна	H_1 верна
Не отвергать H_0 (статистически недостоверные результаты)	Корректное заключение. H_0 верна и не отвергнута	Ошибка второго рода, или β -ошибка. H_1 верна, но H_0 не была отвергнута
Отвергнуть H_0 (статистически достоверные результаты)	Ошибка первого рода, или α -ошибка: H_0 верна, но была отвергнута	Корректное заключение: H_1 верна, H_0 была отвергнута

Статистическая сила исследования (*power*) определяется как вероятность опровержения нулевой гипотезы и констатации различий между обследуемыми группами в случае действительного существования таких различий. Очевидно, статистическая сила вычисляется как $1 - \beta$. Например, если в качестве заданного уровня ошибки второго рода принять 10 % ($\beta = 0,10$), что означает невозможность в 10 % случаев отвергнуть нулевую гипотезу при истинном существовании различий между группами, то статистическая сила исследования составит: $1 - 0,10 = 0,90$. Эта цифра означает, что вероятность выявления различий между группами в случае действительного существования таких различий составит 90 %. Как правило, в качестве

максимального допустимого уровня ошибки второго рода принимают величину, равную 0,2, что соответствует статистической силе исследования 0,8 (т. е. 80 %).

На этапе планирования исследования нужно стремиться к тому, чтобы свести к минимуму ошибки обоих родов. Однако в силу взаимозависимости этих двух ошибок попытки уменьшить ошибку одного рода приводят к потенциальному увеличению ошибки другого рода. Уменьшение ошибки одного рода при неизменном уровне ошибки другого рода представляется возможным лишь за счет увеличения численности выборки.

При расчете оптимальной численности выборки необходимо определить:

- каковы желаемые значения ошибки первого и второго рода;
- какова выраженность эффекта, которую желательно зафиксировать;
- какова предполагаемая доля лиц в контрольной группе, подвергшихся изучаемому воздействию (в исследованиях «случай — контроль»);
- какова предполагаемая доля лиц в популяции, имеющих изучаемый эффект (для когортных или интервенционных исследований).

Оценка доли лиц, которые подверглись влиянию изучаемого воздействия или имеют проявления эффекта этого воздействия, обычно основана на опубликованных ранее данных. Предполагаемая выраженность эффекта может быть также оценена на основе ранее публиковавшихся данных или взята произвольно, с учетом приемлемого, с точки зрения исследователя, минимального значения, свидетельствующего о существовании ассоциации. После того как указанные параметры заданы, возможен расчет численности выборки (Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Woodward M., 2014).

Расчеты необходимой численности выборки можно провести с помощью ряда прикладных статистических программ: EpiInfo (Dean A. G. [et al.], 2011) (<http://www.cdc.gov/epiinfo>), G*Power (Faul F. [et al.], 2007) (<http://www.gpower.hhu.de/en.html>), WINPEPI (Abramson J. H., 2011) (<http://www.brixtonhealth.com/pepi4windows.html>).

При определении численности выборки, необходимой для сравнения независимых средних показателей в двух сопоставляемых группах, пользуются следующими расчетами:

$$n_1 = kn_0,$$
$$n_0 = \frac{(Z_{\alpha/2} + z_{\beta})^2 \cdot \sigma^2 \cdot \left(1 + \frac{1}{k}\right)}{(\bar{X}_1 - \bar{X}_0)^2}$$

где k — соотношение числа наблюдений в контрольной n_1 и основной n_0 группах; $Z_{\alpha/2}$ — стандартное нормальное распределение, соответствующее заданному уровню α ; z_β — стандартное нормальное распределение, соответствующее заданному уровню β (Cohen B. H., Lea R. B., 2004; Chow S.-C. [et al.], 2008). В простейшем случае каждому наблюдению в основной группе соответствует одно наблюдение в контрольной и число наблюдений в основной и контрольной группах одинаково ($k = 1$). Если при этом учесть, что выраженность эффекта (d Коэна) равна:

$$d = \frac{\bar{X}_1 - \bar{X}_0}{\sigma},$$

то формула будет иметь вид:

$$n = 2 \cdot \left(\frac{(Z_{\alpha/2} + z_\beta)}{d} \right)^2.$$

В табл. 26 приведены значения численности выборки (суммарно для двух групп) при соотношении числа наблюдений в основной и контрольной группах 1 : 1 для разных значений ошибок первого и второго рода, а также разной выраженности эффекта.

Предположим, планируется исследование, направленное на выявление выраженного эффекта ($d > 0,8$) какого-либо предполагаемого воздействия. При этом должны быть сформированы две группы: основная и контрольная в соотношении 1 : 1 (одно контрольное наблюдение для каждого «случая»). Пусть в качестве порогового уровня ошибки первого и второго рода приняты $\alpha = 0,05$; $\beta = 0,2$ соответственно. Минимальное число наблюдений (суммарно для обеих групп), которое позволит обеспечить заданные условия, составляет 51 (см. табл. 26). Поскольку речь идет о двух группах с равной численностью, это значение нужно увеличить до ближайшего четного числа, т. е. 52, тогда в каждой группе будет по 26 наблюдений. Разумеется, при необходимости нужно сделать поправку на возможные «потери» наблюдений во время исследования, соответствующим образом увеличив численность групп.

Требумая численность выборки (суммарно для двух групп) для выявления различий двух независимых средних показателей. Двусторонний критерий. Соотношение наблюдений в двух группах 1 : 1

Статистическая сила (1 - β)	Численность выборки									
	$\alpha = 0,05$			$\alpha = 0,01$			$\alpha = 0,001$			
	Малая выраженность эффекта ($d = 0,2$)	Средняя выраженность эффекта ($d = 0,5$)	Сильная выраженность эффекта ($d = 0,8$)	Малая выраженность эффекта ($d = 0,2$)	Средняя выраженность эффекта ($d = 0,5$)	Сильная выраженность эффекта ($d = 0,8$)	Малая выраженность эффекта ($d = 0,2$)	Средняя выраженность эффекта ($d = 0,5$)	Сильная выраженность эффекта ($d = 0,8$)	Сильная выраженность эффекта ($d = 0,8$)
0,60	492	80	33	804	131	53	1261	206	84	
0,65	552	90	36	880	144	58	1357	222	90	
0,70	619	101	41	964	157	63	1461	238	96	
0,75	696	113	45	1060	172	69	1578	257	104	
0,80	787	128	51	1171	190	76	1713	279	112	
0,85	900	146	58	1308	212	85	1878	305	122	
0,90	1053	170	68	1491	241	96	2096	340	136	
0,95	1301	210	83	1785	288	115	2441	395	158	

Для расчета численности выборки, необходимой в исследованиях «случай – контроль» с использованием дискретных данных, формула будет иметь вид:

$$n = \frac{(p_0 \cdot q_0 + p_1 \cdot q_1) \cdot \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta} \right)^2}{(p_1 - p_0)^2},$$

где n – число наблюдений в каждой группе; p_1 – доля лиц, подвергшихся изучаемому воздействию среди «случаев»; p_0 – доля лиц, подвергшихся изучаемому воздействию среди контрольных наблюдений; $q_1 = 1 - p_1$; $q_0 = 1 - p_0$; $Z_{\alpha/2}$ – значение стандартного нормального распределения, соответствующее заданному уровню α ; z_{β} – значение стандартного нормального распределения, соответствующее заданному уровню β (Cohen B. H., Lea R. B., 2004; Chow S.-C. [et al.], 2008). Требуемая численность выборки для проведения подобных исследований представлена в табл. 27.

Таблица 27

Численность выборки (суммарно для двух групп) для выявления различий в двух независимых пропорциях (p_1 и p_0). Двусторонний критерий. Пропорция $p_0 = 0,5$. Соотношение числа наблюдений в двух группах 1 : 1

Статистическая сила ($1 - \beta$)	Пропорция p_1 (или $1 - p_1$)							
	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90
$\alpha = 0,05$								
0,60	1954	484	212	117	73	49	35	25
0,65	2194	544	238	131	82	55	38	28
0,70	2461	610	267	147	91	61	43	31
0,75	2767	685	300	165	102	68	48	35
0,80	3129	775	339	186	115	77	54	39
0,85	3579	886	387	212	132	88	61	44
0,90	4188	1036	452	248	153	102	71	51
0,95	5179	1281	559	306	189	125	87	62
$\alpha = 0,01$								
0,60	3193	792	347	191	119	80	57	41
0,65	3497	867	380	209	130	87	62	45
0,70	3833	950	416	229	143	96	67	49
0,75	4213	1044	457	251	156	105	73	53
0,80	4657	1153	504	277	172	115	81	58

Статистическая сила $(1 - \beta)$	Пропорция p_i (или $1 - p_i$)							
	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90
$\alpha = 0,01$								
0,90	5932	1468	641	352	218	145	102	73
0,95	7101	1757	767	421	260	173	121	86
$\alpha = 0,001$								
0,60	5010	1242	545	301	187	126	89	65
0,65	5390	1336	586	323	201	135	95	69
0,70	5805	1439	630	347	216	145	102	74
0,75	6270	1554	680	375	233	156	110	80
0,80	6809	1687	738	406	253	169	119	86
0,85	7466	1849	809	445	276	185	130	94
0,90	8335	2064	902	496	308	205	144	104
0,95	9711	2403	1050	576	357	238	166	119

В исследованиях популяционного среза расчет численности выборки имеет специфические особенности. Прежде всего, требуется информация о предполагаемой распространенности изучаемого состояния в популяции, т. е. предполагаемой пропорции p . Такая информация не всегда доступна, поэтому чаще всего приходится ориентироваться на «худший сценарий», когда предполагается, что пропорция равна 0,5 (*erred to 0,5*). При иных вариантах частота распространенности в популяции лиц, относящихся к одной из двух групп, — имеющих распознаваемое состояние или нет, — так или иначе будет превышать 0,5. Далее необходимо определиться с тем, на какую минимальную выраженность эффекта рассчитано исследование, т. е. с какой точностью допускается погрешность в оценке распространенности изучаемого состояния. Чем меньше такая погрешность, тем более строгие требования предъявляются к исследованию и к численности выборки. Это показатель **ошибки случайной выборки e** (*margin of error*). Необходимо определиться с допустимой вероятностью нулевой гипотезы α и соответствующим ему значением плотности нормального распределения z_{α} . Наконец, требуются сведения о численности популяции, из которой будет формироваться выборка N . При малой численности целевой популяции N будет иметь принципиальное значение, однако если численность целевой популяции превышает 400, на расчеты численности выборки N уже не оказывает существенного влияния.

С учетом сказанного расчет численности выборки проводится по формуле (Franklin S., Walker C., 2010):

$$n = \frac{z^2 \cdot p \cdot (1 - p)}{e^2 + \frac{z^2 \cdot p \cdot (1 - p)}{N}}$$

Приведены значения численности выборки, необходимые при исследованиях популяционных срезов (табл. 28). Предположим, планируется проведение исследования популяционного среза, направленного на выявление распространенности определенного состояния в популяции. Поскольку сведения о предполагаемой распространенности исходно отсутствуют, следует принять «худший сценарий» и $p = 0,5$. Пусть допустимая ошибка случайной выборки составляет 0,05: это означает, что нас вполне устроит погрешность оценки распространенности изучаемого состояния в пределах $\pm 5\%$. Наконец, пусть вероятность ошибки первого рода α составляет 0,05. Заданным условиям будет удовлетворять численность выборки 385 наблюдений (см. табл. 28). Эта численность может быть меньше, если будет принята в качестве допустимой большая погрешность (например, при погрешности 0,1 требуемая численность выборки составит 97 наблюдений).

Таблица 28

Расчетная численность выборки для исследования популяционного среза (простая случайная выборка) при предполагаемой пропорции $p = 0,5$ и различных значениях ошибки случайной выборки e и ошибки первого рода α

Ошибка случайной выборки e	Ошибка первого рода α , %		
	5	1	0,1
0,025	1537	2654	4332
0,050	385	664	1083
0,075	171	295	482
0,100	97	166	271
0,125	62	107	174
0,150	43	74	121
0,175	32	55	89
0,200	25	42	68

В случаях корреляционных исследований численность выборки непосредственно связана с коэффициентом корреляции, который представляет собой выраженность эффекта. В табл. 29 представлены сведения о требуемой численности выборки в случаях проведения корреляционных исследований. Так, если планируется проведение

исследования с корреляционным дизайном, направленное на выявление сильно выраженного эффекта ($\rho > 0,5$), то при допустимых $\alpha = 0,05$ и $\beta = 0,2$ численность выборки должна быть не менее 26 наблюдений. Для констатации менее выраженного эффекта требуется увеличение численности выборки: например, для констатации средней выраженности эффекта потребуется минимум 81 наблюдение.

Таблица 29

Требуемая численность выборки для корреляционного исследования при различных значениях статистической силы, выраженности эффекта и вероятности ошибки (двусторонний критерий)

Статистическая сила ($1 - \beta$)	Выраженность эффекта								
	малая ($\rho = 0,1$)	средняя ($\rho = 0,3$)	сильная ($\rho = 0,5$)	малая ($\rho = 0,1$)	средняя ($\rho = 0,3$)	сильная ($\rho = 0,5$)	малая ($\rho = 0,1$)	средняя ($\rho = 0,3$)	сильная ($\rho = 0,5$)
	$\alpha = 0,05$			$\alpha = 0,01$			$\alpha = 0,001$		
	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	
	487	52	17	796	84	27	1249	132	43
	546	58	19	871	92	30	1343	142	46
	613	64	21	955	101	32	1446	153	49
	689	72	23	1049	110	35	1562	164	53
	779	81	26	1160	121	38	1696	178	57
	891	93	29	1295	135	43	1859	195	62
	1042	108	34	1476	154	48	2075	217	68
	1288	133	41	1767	183	57	2417	252	79

Вопросы для самоконтроля

1. Каковы принципы суммирования и представления данных, полученных в ходе исследования?
2. В чем суть оценки влияния случая на результаты исследования? Что такое нулевая гипотеза и каковы подходы к оценке ее вероятности?
3. В чем смысл оценки доверительных интервалов?
4. Что такое выраженность эффекта?
5. Каковы подходы к оценке требуемой численности выборки?

ОЦЕНКА ВАЛИДНОСТИ ИССЛЕДОВАНИЯ

В предыдущих разделах понятие «валидность» рассматривалось применительно к измерению с помощью определенной методики. Однако валидность может иметь и более широкую трактовку, а именно применительно к тем выводам, которые можно в целом сделать на основании исследования. В таком случае речь идет о **валидности исследования** (*research validity*) (Shadish W. R. [et al.], 2001; Barker C. [et al.], 2002).

В соответствии с классическими подходами к оценке валидности исследования, ее можно разделить на внутреннюю и внешнюю (Campbell D. T., Stanley J. C., 1963) (рис. 52).

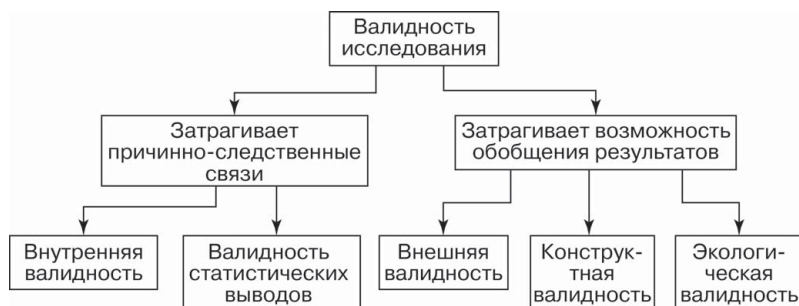


Рис. 52. Важнейшие виды валидности исследования

Внутренняя валидность (*internal validity*) рассматривает вопрос: насколько результаты исследования позволяют говорить о наличии причинно-следственных связей? Иными словами: связаны ли на самом деле изменения зависимой переменной с влиянием изменений независимой переменной?

Внешняя валидность (*external validity*) призвана ответить на вопрос: в какой мере результаты исследования можно распространять (обобщать) применительно к иному моменту времени, иным условиям, пациентам, ситуациям и т. п. ? Близким по смыслу (хотя и не

тождественным) является понятие «**экологическая валидность**» (*ecological validity*), которая изучает, насколько искусственными по сравнению с реальными жизненными условиями были измерения и процедуры исследования. В дополнение к внутренней валидности было предложено рассматривать **валидность статистического вывода** (*statistical conclusion validity*), которая также призвана интерпретировать ковариацию независимой и зависимой переменных, а в дополнение к внешней валидности предложено рассматривать **конструктивную валидность** (*construct validity*), направленную на оценку возможности обобщения результатов исследования в отношении конструкторов более высокого порядка с учетом того, насколько удовлетворительно измерен изучаемый конструктор методиками исследования (Cook T. D., Campbell D. T., 1979; Shadish W. R. [et al.], 2001).

Принципиально важно, что возрастание внутренней валидности исследования нередко сопровождается снижением внешней, и наоборот. Так, при проведении лабораторных исследований можно обеспечить строгие условия эксперимента, повысив тем самым внутреннюю валидность, однако условия исследования будут избыточно искусственными, и возможность обобщения полученных результатов на популяцию в целом (внешняя валидность) будет уменьшаться. С другой стороны, исследования в полевых условиях будут максимально приближены к естественным условиям (высокая внешняя валидность), однако нередко в ущерб внутренней валидности за счет влияния ряда неконтролируемых факторов.

Таким образом, любые дизайны и методики исследования не могут быть идеальными, но в каждом случае возможен систематический анализ возможных причин и последствий такого несовершенства, которые именуются **угрозами валидности** (*threats to validity*) (Shadish W. R. [et al.], 2001). Планируя и осуществляя исследование, автор должен стремиться к тому, чтобы свести к минимуму угрозы для валидности, а впоследствии уже на этапе интерпретации и обобщения полученных результатов, ему нужно критически оценить, какова вероятность наличия таких угроз и в какой мере они могли отразиться на результатах.

5.1. Ассоциации и причинно-следственные связи

Выявленная ассоциация между изучаемыми в ходе исследования признаками еще не дает оснований говорить о наличии причинно-следственных связей между ними. Эта проблема особо актуальна в исследованиях с неэкспериментальным дизайном. «Золотое» правило гласит: **корреляция не эквивалентна причинности** (*correlation does not equal causation*) (Barker C. [et al.], 2002). Корреляционные связи

между изучаемыми признаками могут лишь наводить на мысль о наличии причинно-следственных связей, но еще не позволяют с уверенностью делать подобное утверждение. В клинической психологии, как и в медицинских науках, исследователь имеет дело с вероятностными, а не детерминистскими закономерностями. Так, если мы говорим о том, что неблагоприятные социально-демографические характеристики могут повышать риск аффективных расстройств у пациента, это еще не означает, что речь идет о непосредственных причинно-следственных связях. Аналогично, если делается вывод о том, что интервенционное воздействие сопровождается положительным эффектом, то подразумевается, что интервенция создает условия, когда вероятность позитивных сдвигов повышается, однако это не значит, что положительные сдвиги неизбежны в каждом случае.

Причинно-следственная связь между анализируемыми в исследовании показателями теоретически возможна при соблюдении ряда условий (Haynes S. N., O'Brien W. H., 2000). Во-первых, между двумя показателями должна иметься ковариация — математически доказанная зависимость. Поскольку суждение о наличии такой зависимости основано на статистических критериях, требуется критическая оценка валидности статистических выводов. Вопросы, связанные с ее обеспечением, рассматривались в главе 4, однако, резюмируя сказанное, можно указать следующие важнейшие виды угроз валидности статистических выводов, требующих критической оценки:

— *Низкая статистическая сила исследования.* Это обстоятельство может приводить к некорректному выводу об отсутствии связи показателей, тогда как фактически такая связь имеется.

— *Неверные допущения, приписываемые статистическим тестам.* Выраженность реального эффекта может быть как заниженной, так и завышенной.

— *Повторные измерения («забрасывание сетей», fishing) и связанная с этим ошибка.* Повторные многочисленные измерения, направленные на выявление статистически достоверных ассоциаций, могут искусственно завышать вероятность их выявления, если не вносится соответствующая поправка.

— *Ненадежность измерений.* Ошибка, закрадывающаяся в измерения, снижает выраженность ассоциации между двумя изучаемыми переменными и повышает или снижает выраженность ассоциаций трех переменных и более.

— *Сужение диапазона анализируемой переменной.* Такое искусственное сужение способно снизить выраженность ассоциации между этой и другой переменной.

— *Ненадежность внедряемого лечения.* Если планируется внедрение лечения в стандартной форме, а эффективность этого лечения

изучалась при его использовании в ограниченных масштабах у части испытуемых, эффективность полномасштабного лечения может быть недооценена.

— *Влияние внешних переменных в условиях эксперимента.* Условия, связанные с организацией экспериментального исследования, могут усиливать влияние ошибки, что делает более затруднительным выявление эффекта.

— *Гетерогенность изучаемых единиц наблюдения.* Повышенная вариабельность оцениваемой зависимой переменной увеличивает ошибку и затрудняет выявление связи.

— *Неточная оценка выраженности эффекта.* Неправильно выбранные статистические методы могут систематически занижать или завышать выраженность эффекта (Shadish W. R. [et al.], 2001).

Во-вторых, суждение о причинно-следственных связях не должно противоречить законам логики. Исследователя подстерегают многочисленные риски, например логические заблуждения при обсуждении вопроса о причинно-следственной связи между изучаемыми явлениями (Damer T. E., 2009). Опираясь на классификацию Аристотеля, Г. Селье выделил 8 основных типов **логических заблуждений**:

— *От приводящего:* когда приводящий факт принимается за существенный. Например, данный гормон белый, вывод — все гормоны белые.

— *Secundum quid* («следует из общего» — *лат.*) — ошибочное распространение общего правила на частный случай и наоборот, без учета изменяющихся обстоятельств (например, если стресс вызывает дегенерацию вилочковой железы, то у крысы с удаленными надпочечниками стресс должен повлечь за собой такую же дегенерацию; на самом деле стресс воздействует на вилочковую железу через надпочечники, поэтому его действие блокируется при их удалении).

— Неверные заключения — особый способ убеждения за счет отвлечения внимания на какой-либо посторонний факт (например, нападки на приверженцев теории вместо доказательства ее сложности).

— Предрешение основания — выдвигание в качестве доказательства того, что само требует доказательства; выбор таких посылок, которые заранее предопределяют вывод еще не доказанных посылок (например, отрицание довода на том основании, что он «ненаучен», что также нужно доказать).

— Ошибка следования — рассуждение от последовавшего события к его условию (например, удаление паразитовидной железы вызывает судороги, следовательно, судороги указывают на отсутствие гормона паразитовидной железы).

— *Non sequitur* («не следует» — *лат.*) — обоснование заключения исходя из недостаточного или ложного факта (например, холод явля-

ется стрессором, холод вызывает дрожь, следовательно, дрожь является неспецифическим проявлением действия всех стрессоров).

— *Post hoc ergo propter hoc* («после этого — значит, по причине этого» — *лат.*) — если изменение появляется после воздействия какого-либо фактора, оно должно быть обусловлено этим фактором (например, если эксперимент удался весной, но не удался осенью, его успех зависит от смены времен года).

— Ошибка совмещения множества вопросов — неправомерное объединение нескольких вопросов в один (например, почему кортизон является самым полезным из всех кортикоидов? В данном случае утверждение о полезности кортизона некорректно вводится как доказанный факт) (Селье Г., 1987).

В-третьих, обоснование возможного наличия причинно-следственных связей предполагает отсутствие альтернативных объяснений выявленной ассоциации. Однако на практике такие альтернативные объяснения часто могут существовать и всегда требуют специального рассмотрения при обсуждении и интерпретации полученных результатов.

Пример

Выявлена ассоциация между признаками А и В. Так, может быть выявлена статистически достоверная связь (положительная корреляция) между уровнем эмпатии психотерапевта (признак А) и эффектом психотерапии (признак В). При этом возможны несколько допущений (инференций) относительно причинно-следственных связей, которые схематично отражены на рис. 53. Во-первых, признак А может быть причинно значимым, а признак В — следствием (более развитая эмпатия психотерапевта способствует выздоровлению клиента, см. вариант 1). С другой стороны, не исключено, что признак В является причиной, а признак А — следствием (выздоровление пациента на фоне лечения вызывает большую эмпатию со стороны психотерапевта, см. вариант 2). Часто встречается и такая ситуация, когда выявляемая ковариация признаков А и В целиком объясняется влиянием неучтенного признака С.

Внушаемость пациента может влиять как на эффективность психотерапии, так и на формирование большей эмпатии со стороны психотерапевта. Такой третий фактор С (вариант 3) выступает в роли конкурирующего при объяснении причинности и существенно снижает внутреннюю валидность исследования. Фактор С в такой ситуации является искажающим и подробно обсуждается в соответствующем разделе. Еще одно возможное объяснение выявленным ассоциациям состоит в допущении того, что фактор А влияет на фактор В, но не напрямую, а опосредованно, через фактор D. В таком случае D обозначается как *медиатор*, так как выполняет посредническую миссию (см. вариант 4). Эмпатия психотерапевта может способствовать

более глубокому самоанализу пациента. Это, в свою очередь, способствует лучшему исходу проводимого лечения.

Наконец, не исключена и такая ситуация, когда выраженность ассоциации факторов А и В зависит от воздействия фактора Е и его уровня. Фактор Е в этом случае выступает в роли *модератора*, как своеобразный клапан, фильтр (см. вариант 5). В роли модератора могут выступать форма психологического нарушения у пациента (тревога или депрессия), социально-демографические характеристики. Так, если эмпатия психотерапевта лучше реализуется при работе с пациентами-женщинами, можно сказать, что причинно-следственная

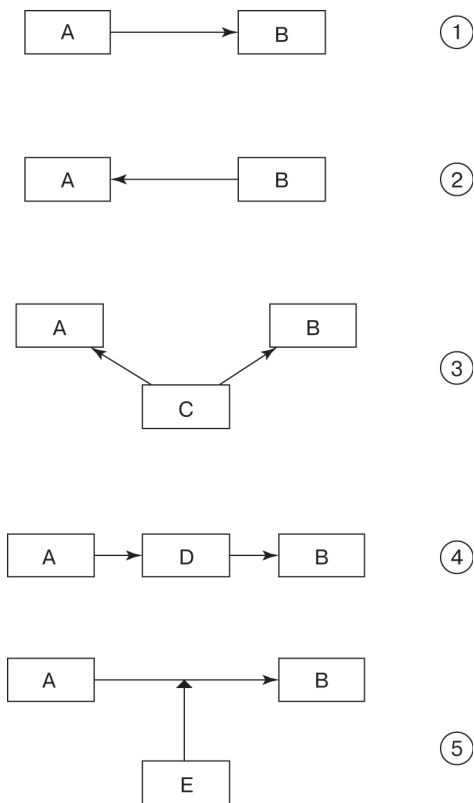


Рис. 53. Различные варианты причинно-следственных связей в случае выявленной ассоциации признаков А и В

связь между эмпатией и эффективностью терапии модифицируется полом пациента, пол выступает в качестве признака-модератора.

Наличие фактора-медиатора и фактора-модератора в выявленной ассоциации не столь критично для внутренней валидности исследования, как наличие искажающего фактора, так как в целом присутствие медиатора или модератора не дискредитирует выявленную ассоциацию, однако такую возможность следует рассматривать и обсуждать. Кроме того, имеются и специальные статистические методы, направленные на оценку возможного присутствия факторов-медиаторов и модераторов.

Есть еще одно возможное объяснение, которое часто не учитывается, но требует рассмотрения. Ассоциация может отражать **концептуальное искажение** (*conceptual confound*): смысл переменных А и В может быть близок. Так, эмпатия психотерапевта (особенно если ее оценивает пациент) может в определенной степени отражать эффект психотерапии, тогда выявленная ассоциация отражает связь разных сторон эффективности психотерапии.

Итак, при планировании дизайна исследования требуется предусмотреть сбор данных в таком объеме, который позволит проанализировать влияние третьей переменной, включая эффект медиатора и модератора, концептуального искажения, чтобы прийти к обоснованному заключению о наличии связи между изучаемыми показателями.

Выделяют ряд позитивных аргументов, подтверждающих возможность причинно-следственной связи между изучаемыми признаками. Эти аргументы были предложены Остином Брэдфордом Хиллом (Austin Bradford Hill) и известны в литературе как критерии Хилла. К числу таких аргументов относятся:

- выраженность ассоциации;
- биологическая обоснованность гипотезы;
- согласованность полученных результатов с результатами других исследований;
- последовательность событий во времени;
- зависимость эффекта от выраженности воздействия (зависимость типа «доза — эффект») (Hill A. B., 1965).

Для анализа гипотезы представляется важным учет выраженности выявленной ассоциации. Тогда допустимо сделать вывод о том, в какой степени само по себе изучаемое воздействие может повлиять на риск развития заболевания, а следовательно, какова вероятность того, что между воздействием и заболеванием существует причинно-следственная связь. Чем сильнее выявленная ассоциация, т. е. чем больше увеличение (или уменьшение) риска, тем меньше вероятность того, что эффект связан с воздействием какого-либо неучтенного постороннего фактора.

Пример

Исследование эпидемиологии синдрома внезапной смерти младенцев выявило, что в Скандинавских странах риск смерти ребенка выше в 4,1 раза, если мать курила во время беременности. Для того чтобы объяснить такое ошутимое возрастание риска, не связанное с причинным влиянием курения матери, следует допустить существование какого-то иного, неучтенного фактора, который во много раз чаще наблюдается среди курящих женщин по сравнению с некурящими беременными и который сам по себе существенно влияет на риск синдрома внезапной смерти младенцев.

Приведенные рассуждения не означают, что ассоциация малой выраженности не может отражать причинно-следственную связь между явлениями. Можно лишь говорить о том, что в подобных случаях затруднительно исключить альтернативные объяснения.

Поверить в реальность причинно-следственной связи, стоящей за выявленной ассоциацией, становится гораздо легче, если известны конкретные биологические механизмы, объясняющие, каким образом анализируемое патологическое воздействие может повлиять на риск заболевания.

Пример

Выявленная ассоциация между симптомами депрессии у матери во время беременности и риском задержки внутриутробного развития плода может быть обоснована, если принять во внимание повышенную вероятность патологии фетоплацентарного кровообращения, изменений гормонального профиля, нарушений питания женщины на фоне депрессии. В свою очередь, эти факторы оказывают негативное влияние на рост и развитие плода (Кельмансон И. А., 2015).

Поскольку вывод о биологической обоснованности зависит от уровня медицинских и клиничко-психологических знаний на конкретном этапе, отсутствие достоверных сведений о биологической обоснованности гипотезы не исключает реальной причинно-следственной связи.

Наиболее убедительными выявленные ассоциации становятся тогда, когда сходные результаты дают многочисленные исследования, проведенные различными авторами, в различных условиях, в различных по своим демографическим и социальным характеристикам популяциях, с использованием различных методик. Отдельное исследование может быть подвержено влиянию неучтенной систематической ошибки или факторов, искажающих результаты, но трудно предположить, что многочисленные исследования могут испытывать на себе постоянное влияние подобных обстоятельств. Итак, согласованность с результатами других исследований является одним из наиболее веских аргументов в пользу реального существования при-

чинно-следственной связи. Примером могут служить данные о связи курения матери во время беременности с повышенным риском нарушений поведения детей раннего возраста: подобная ассоциация была выявлена в ходе многочисленных исследований, проведенных в большинстве промышленно развитых стран мира. Напротив, отсутствие согласованности результатов должно сильно настораживать при попытке обосновать причинно-следственные связи.

Логично также предположить, что при наличии реальной причинно-следственной связи воздействие должно предшествовать результату, а временной промежуток, разделяющий эти события, должен соответствовать представлениям о биологических механизмах, лежащих в основе этой связи. Иными словами, предполагаемое патогенное воздействие должно предшествовать заболеванию, которое рассматривается как его следствие, а предполагаемое лечебное мероприятие — предшествовать по времени тому позитивному эффекту, который рассматривается как его результат. К сожалению, часто представляется затруднительным точно установить такую последовательность событий. Характеристики многих потенциальных причинных факторов могут существенно меняться, как только у пациента проявятся первые симптомы заболевания. В частности, это касается уровня физической активности, особенностей диеты. Подобные изменения могут быть как преднамеренными, так и связанными с непосредственным течением заболевания. Аргументация в пользу того, что анализируемое воздействие реально предшествует развитию того или иного эффекта, наиболее убедительна для проспективных когортных или интервенционных (экспериментальных) исследований. Однако даже в подобных ситуациях могут возникнуть трудности в интерпретации результатов, если временной промежуток между воздействием и анализируемым эффектом достаточно короткий.

В качестве довода в пользу причинности традиционно рассматривается существование зависимости по типу «доза — эффект», т. е. наличие градиента риска, связанного с уровнем воздействия.

Пример

Частота нарушений дыхания во время сна выше у детей первого года жизни, матери которых выкуривали во время беременности 10 и более сигарет в сутки, чем у детей, матери которых выкуривали менее 10 сигарет в сутки. В свою очередь, частота нарушений дыхания во время сна выше у тех детей первого года жизни, матери которых во время беременности выкуривали менее 10 сигарет в сутки, чем у детей, родившихся у некурящих матерей. Это обстоятельство может служить дополнительным аргументом в пользу существования причинно-следственной связи между курением матери во время беременности и риском нарушений дыхания во время сна у ребенка первого года жизни.

Трудность использования данного критерия состоит в том, что, с одной стороны, существование зависимости «доза — эффект» не означает причинно-следственного характера выявленной ассоциации, так как она может отражать воздействие неучтенного фактора, стоящего за такой ассоциацией и реально обуславливающего риск. С другой стороны, отсутствие подобного градиента также не может рассматриваться как абсолютный аргумент против существования причинно-следственной связи:

- в ряде случаев не представляется возможным четко ранжировать уровень воздействия и выраженности эффекта;

- многие медико-биологические феномены, с которыми сталкиваются клинические психологи, подчиняются принципу «все или ничего», т. е. воздействие должно достигнуть определенного уровня, чтобы вызвать эффект.

Очевидно, выявляемые ассоциации требуют к себе весьма критичного отношения. Следует учитывать все имеющиеся аргументы и с осторожностью выносить суждения о реальной причинно-следственной связи анализируемых явлений.

5.1.1. Угрозы внутренней валидности

5.1.1.1. Наличие систематической ошибки

Систематическая ошибка (*bias*) исследования — это фактор, который приводит к неверной оценке ассоциации между изучаемым воздействием и эффектом. Даже самое строгое планирование условий исследования не избавляет от влияния ошибок, связанных с принципом отбора пациентов, сбора, передачи и интерпретации информации. Вот почему в ходе оценки полученных результатов необходимо критически анализировать возможность того, что выявленные ассоциации связаны с влиянием систематической ошибки, и дать оценку вероятности такого объяснения. В отличие от вероятности случайности ассоциации, которая может получить количественную оценку, эффект влияния систематической ошибки трудно поддается оценке, а нередко может просто оказаться неучтенным. В этой связи ключевым моментом в планировании любого исследования является создание таких условий, в которых любое возможное влияние систематической ошибки было бы заранее принято во внимание и сведено к минимуму. Однако даже при соблюдении этих условий на этапе планирования, на заключительном этапе необходимо оценить: не могла ли систематическая ошибка все-таки закрасться в исследование, каков вероятный источник этой ошибки, в каком направлении она могла повлиять на полученные результаты?

Имеются многочисленные классификации систематических ошибок, однако принципиально можно выделить два вида, в рамках которых есть отдельные варианты.

Первый вид можно обозначить как **ошибку отбора** (*selection bias*), источник которой — принцип отбора субъектов для участия в исследовании.

Второй вид обозначается как **обсервационная, или информационная, ошибка** (*observation or information bias*), и данный вид систематической ошибки связан с оценкой информации, касающейся патогенного воздействия и его эффекта (Shadish W. R. [et al.], 2001; Rossi P. H. [et al.], 2004; Rothman K. J. [et al.], 2008; Straus S. E., 2011).

Ошибка отбора. Отбор индивидуумов для участия в большинстве аналитических исследований проводится с учетом имевшего место патогенного воздействия (в когортных исследованиях) либо имеющегося эффекта этого воздействия, например развившегося заболевания (в исследованиях типа «случай — контроль»). Принципиально важно, чтобы на этапе отбора пациентов изучаемое воздействие и изучаемый эффект этого воздействия не учитывались одновременно и не могли одновременно влиять на принятие решения об отборе пациентов для участия в исследовании. Несоблюдение этого условия является источником ошибки отбора. Ошибка отбора возможна, если в исследовании «случай — контроль» отбор испытуемых каким-то образом связан с имевшим место патогенным воздействием, а в когортных исследованиях — с установленным фактом развития заболевания. Ошибка отбора представляет наибольшую опасность для исследований «случай — контроль» и ретроспективных когортных исследований, так как к началу тех и других уже имеется информация о патогенном воздействии и развитии заболевания. Напротив, ошибка отбора маловероятна при осуществлении проспективных когортных исследований, когда на начальной стадии имеется лишь информация о воздействии, но отсутствуют сведения о его эффекте (Кельмансон И. А., 2002).

Ошибку отбора могут спровоцировать многочисленные обстоятельства, влияющие на подход к отбору пациентов для участия в исследовании:

- различия в медицинском наблюдении за отдельными лицами;
- частота обращений пациентов за медицинской и психологической помощью;
- качество диагностики и т. п.

Кроме того, в исследованиях типа «случай — контроль» причиной ошибки отбора могут быть различия в высказывании согласия участвовать в исследовании лиц, относящихся к двум сопоставляемым группам. Например, семьи, характеризуемые низким социальным

статусом, по ряду причин могут с большей вероятностью отказаться от участия в исследовании в качестве контрольных наблюдений, если целью такого исследования является изучение ассоциации между социальным неблагополучием и риском какого-либо заболевания.

Обсервационная (информационная) ошибка возникает в связи с систематическими различиями в способах получения информации об имевшемся предполагаемом патогенном воздействии или развитии заболевания в сопоставляемых группах. Если собранные данные являются неполными или неточными и в сопоставляемых группах эта неточность варьирует, не исключена возможность выявления ложной ассоциации. В зависимости от источника происхождения различают несколько вариантов обсервационных ошибок:

Ошибка воспроизведения (recall bias) возникает:

— когда лица, страдающие анализируемым заболеванием, способны воспроизвести информацию о предполагаемом патогенном воздействии в предшествующий период иначе, нежели лица, не страдающие этим заболеванием. Такая ситуация особо вероятна в исследованиях «случай — контроль»;

— когда лица, подвергшиеся изучаемому воздействию, способны предоставить информацию о потенциальных последствиях этого воздействия иным образом, чем те обследуемые, которые не испытали на себе патогенного воздействия (в проспективных когортных исследованиях).

Естественно, сказанное относится не только к самим обследуемым, но и к их окружению: родителям, другим родственникам, медицинскому персоналу. Ошибка воспроизведения может приводить как к недооценке, так и к переоценке выраженности предполагаемой ассоциации в зависимости от того, как наличие патологического состояния у обследуемых лиц влияет на получаемую информацию.

Ошибка интервьюера (interviewer bias) возникает в связи с систематическими различиями в методике сбора, регистрации и интерпретации интервьюерами информации, полученной у отдельных участников исследования независимо от дизайна исследований:

— В исследованиях «случай — контроль» возникает особая опасность: осведомленность интервьюера о наличии или отсутствии заболевания у обследуемых может способствовать большей склонности к поиску и обнаружению предшествующих патогенных воздействий у лиц, страдающих заболеванием.

— В ретроспективных когортных исследованиях существует та же опасность осведомленности интервьюера и опасность информационной ошибки в форме ошибки интервьюера.

— В интервенционных исследованиях может наблюдаться ошибка интервьюера, если не используется плацебо-контроль и слепой метод.

Источником систематической информационной ошибки в когортных исследованиях может послужить фактор **потери обследуемых** (*loss to follow-up*) в ходе динамического наблюдения, особенно если процент таких потерь существенно различается в зависимости от отношения обследуемых к изучаемому воздействию и его эффекту.

Еще одним вариантом информационной ошибки является **неверная классификация** (*misclassification*), при которой неверно оценивается факт имевшегося патологического воздействия или его эффекта. Ошибки классификации неизбежно возникают в ходе любого исследования, однако их опасность становится наиболее значимой, если наличие изучаемого патологического состояния у пациента в определенном направлении влияет на качество распознавания предшествовавшего патогенного воздействия, или, напротив, факт патогенного воздействия в определенном направлении влияет на распознавание патологического состояния, связанного с этим воздействием. Такая ситуация именуется **дифференцированной ошибкой**, которая может существенно искажать результаты исследования.

Важнейшим принципом, которым следует руководствоваться для предотвращения систематической ошибки, является тщательное планирование исследования. В некоторых случаях можно усмотреть ошибку в ходе исследования и принять меры к ее минимизации или устранению, однако при наличии ошибки отбора ее устранение может представляться маловероятным. Вот почему стадия планирования исследования является в этом отношении критичной.

Можно отметить несколько ключевых позиций, учет которых позволяет избежать систематических ошибок. В их числе принцип формирования обследуемой выборки, требования к источникам информации, методики сбора и анализа данных. Имеются многочисленные факторы, которые предопределяют, в какой мере подходы к формированию выборки повлияют на снижения вероятности систематической ошибки. Например, использование госпитальных контрольных наблюдений в исследованиях «случай — контроль» позволяет улучшить сопоставимость обследуемых с точки зрения их желания принять участие в исследовании, нивелировать влияние факторов, предопределивших выбор медицинского учреждения, куда были госпитализированы пациенты и где они были обследованы. Указанные обстоятельства позволяют снизить вероятность ошибки, связанной с отказом пациентов от участия в исследовании, избежать ошибки отбора и воспроизведения. Для когортных исследований и клинических испытаний, когда принципиальна возможность следить за судьбами обследуемых в течение заданного промежутка времени, основой отбора нередко является наличие точных сведений о месте жительства и занятости испытуемых. Кроме того, включение в иссле-

дование тех лиц, которые имеют повышенный риск развития изучаемого заболевания, позволяет ожидать большей заинтересованности в сотрудничестве с исследователем и более регулярных контактах.

Во многих аналитических исследованиях методика сбора информации может оказать принципиальное влияние на полученные результаты. С практической точки зрения минимизировать влияние систематической информационной ошибки удастся, если обеспечены:

- создание специального инструментария сбора информации в форме тестов, опросников, стандартных бланков и т. п.;

- разработка такого протокола исследования, который позволит оптимально использовать выбранный инструментарий. Независимо от того, какова методика исследования, она должна последовательно реализовываться в процессе сбора информации в каждой из сопоставляемых групп (Кельмансон И. А., 2002).

Инструментарий. Наиболее существенным резервом снижения влияния систематической ошибки является использование максимально стандартизованных и объективных методик. Если предметом исследования являются показатели артериального давления, информацию обеспечивают данные анкетирования пациентов и данные нескольких измерений, проведенных квалифицированным медицинским персоналом с соблюдением техники измерения; очевидно, второй способ является более объективным и, следовательно, более предпочтительным. Если же источником информации является опрос обследуемых, следует стремиться к максимальной точности формулировок вопросов и однозначности их интерпретации анкетруемыми.

Протокол. Важнейшим подходом к минимизации потенциальной ошибки является стремление к сбору информации «вслепую». Персонал, который осуществляет сбор данных или обследование пациентов, не должен быть осведомлен о том, испытал ли пациент изучаемое воздействие, если проводится проспективное когортное или интервенционное исследование. Персонал не должен быть осведомлен о том, имеется ли у пациента изучаемое заболевание, предполагаемый эффект воздействия, если изучается возможное влияние предшествовавшего патогенного воздействия в ходе исследования «случай — контроль». Более того, весьма желательно, чтобы сами обследуемые лица, по возможности, были минимально осведомлены о том, каков их собственный статус в исследовании, а также о том, какова суть гипотезы исследования. Очевидно, пределы неосведомленности во многом определяются этическими соображениями. Хотя эффективность исследования «вслепую» является доказанной с точки зрения снижения риска систематической информационной ошиб-

ки, полное соблюдение этого принципа не всегда представляется возможным.

Регламентация. Не менее важным подходом, позволяющим снизить потенциальный риск информационной ошибки, является обучение персонала стандартной и строгой методике сбора данных, использование четко регламентированного протокола исследования. Иногда берутся на вооружение специальные приемы, позволяющие заподозрить нарушения протокола, несоблюдение идентичности подхода к сбору информации в двух сопоставляемых группах. Так, возможно сопоставление частоты, с которой лица, относящиеся к двум группам, указывают на наличие у них фиктивных переменных (*dummy variables*), никак не связанных ни с предполагаемым патогенным воздействием, ни с изучаемым заболеванием или, напротив, имеющих с ними доказанную, четко определенную связь.

Пример

При исследовании возможной ассоциации между приемом матерью транквилизаторов во время беременности и риском нарушений сна у ребенка первого года жизни можно попросить обследуемых женщин дать ответ на вопрос: принимали ли они во время беременности, кроме транквилизаторов, иные лекарственные препараты, в частности такие лекарства, для которых доказано отсутствие связи с последующим риском нарушений сна у ребенка. Например, можно попросить ответить на вопрос о приеме ими слабительных препаратов. Если сопоставление частоты употребления лекарственных препаратов во время беременности женщинами, дети которых имели нарушения сна, и женщин, дети которых не имели нарушений сна, выявит более частое использование транквилизаторов у женщин из первой группы, но не выявит различий в использовании слабительных, можно думать о том, что ассоциация действительно является значимой. Напротив, наличие достоверных различий между двумя сопоставляемыми группами в частоте использования и транквилизаторов и слабительных позволит заподозрить, что они связаны с влиянием ошибки воспроизведения или ошибки интервьюера. С другой стороны, в опросник можно было бы включить сведения о курении матери во время беременности — доказанном факторе риска нарушений сна у ребенка первого года жизни (Kelmanson I. A., 2009). Если в ходе исследования одновременно с выявлением ассоциации между использованием матерью транквилизаторов во время беременности и риском нарушений сна у ребенка будет также воспроизведена и доказанная ассоциация между курением матери во время беременности и риском нарушений сна у ребенка, это может послужить дополнительным аргументом в пользу валидности исследования.

Дублирующие вопросы. В опросник включают несколько вопросов, сформулированных по-разному, но принципиально направленных на поиск одной и той же ассоциации. Различия в выявленных

ассоциациях, оцениваемых на основе ответов на дублирующие друг друга вопросы, могут привести на мысль о существовании систематической ошибки. **Учет времени**, потраченного интервьюером на беседу с пациентом, позволяет определить, не тратит ли интервьюер систематически больше или меньше времени на получение информации у представителей той или иной группы, что также может явиться источником систематической ошибки.

Наконец, может быть полезным **включение в опросник шкалы, позволяющей интервьюеру дать собственную субъективную оценку того, насколько конкретный пациент адекватно воспринимает задаваемые ему вопросы и дает на них ответы**. В дальнейшем субъекты с неудовлетворительными оценками качества своих ответов могут быть исключены из исследования или проанализированы отдельно.

Источник информации оказывает существенное влияние и на вероятность систематической ошибки. Сведения об имевшемся патогенном воздействии и о наличии заболевания могут быть получены в результате непосредственного опроса обследуемых лиц, анализа медицинской документации, статистики рождаемости и смертности, а также в результате непосредственного измерения интересующих показателей. Наиболее надежным источником информации следует признать медицинскую документацию, оформленную до начала исследования, как наименее подверженную влиянию субъективных оценок, связанных с целями и задачами исследования, и максимально свободную от влияния информационной ошибки. В то же время такие архивные документы могут не содержать сведений о ряде признаков, непосредственно интересующих исследователя. Более того, степень информационных пробелов может существенно различаться в отдельных сопоставляемых группах наблюдений.

Пример

Анализ архивных документов, относящихся к детям из отдельных категорий риска (недоношенные, маловесные, имеющие врожденные пороки развития и т. п.), убеждает в более пристрастном отношении медицинских работников к активному выявлению отягощающих моментов акушерско-гинекологического анамнеза, чем в случае рождения ребенка без явных клинических признаков патологии (Кельмансон И. А., 2002). Данный пример иллюстрирует необходимость по мере возможности использовать и сопоставлять различные источники информации.

Исследования, основанные на анализе медицинской документации, могут предполагать использование стационарных историй болезни, патологоанатомических протоколов, а также амбулаторных

данных. Сведения, полученные с помощью опросников и анкетирования обследуемых субъектов, могут дополняться и верифицироваться данными анализа медицинской документации. Следует понимать, что потребность в верификации данных о наличии патогенного воздействия или заболевания во многом зависит от самой природы изучаемого заболевания. Все анализируемые воздействия и вероятные исходы этих воздействий должны иметь четкие критерии для их констатации. В случаях распознавания отдельных заболеваний (патологических состояний) нередко говорят о существовании принятого «золотого стандарта», и качество исследования во многом определяется тем, использовался ли этот стандарт для верификации диагноза. Нередко в качестве «золотого стандарта» выступают критерии, разработанные экспертами ВОЗ или иными авторитетными экспертами на основе консенсуса.

Несмотря на усилия избежать систематической ошибки на стадии планирования исследования, всегда существует вероятность ее наличия. Вот почему на стадии анализа и обсуждения полученных результатов требуется дать критическую оценку:

— Какая ошибка и в какой мере могла повлиять на полученные результаты?

— В каком направлении — маскирования истинной ассоциации или ее гиперболизации — систематическая ошибка могла повлиять на полученные результаты?

— Какова возможная степень этого искажения?

Такие оценки не имеют количественного выражения и не могут быть проверены с использованием какого-то статистического критерия: они основаны лишь на здравом смысле и понимании конкретной клинической ситуации, сопряженной с исследованием.

Следует также иметь в виду, что отдельные варианты исследований в разной мере подвержены влиянию систематических ошибок конкретных видов:

— *Исследования «слушай — контроль».* В максимальной степени возможна ошибка, связанная со следующими обстоятельствами:

— сведения о статусе обследуемого в отношении наличия или отсутствия у него заболевания могут влиять на принятие решения в отношении имевшегося патогенного воздействия (ошибка воспроизведения);

— сведения об имевшемся или отсутствовавшем патогенном воздействии могут повлиять на выявление распознаваемого заболевания (ошибка отбора) (Кельмансон И. А., 2002).

— *Когортные исследования.* Особую тревогу вызывает ошибка, связанная с потерями обследуемых в ходе динамического наблюдения.

— *Интервенционные исследования.* Вероятность наблюдательной ошибки зависит во многом от характера сопоставляемых групп, использования плацебо и степени объективности оценки результатов.

В любом исследовании возможна ошибка классификации. При этом принципиально важно следующее:

— *дифференцированная ошибка* может приводить как к занижению, так и к завышению реальной ассоциации;

— *недифференцированная ошибка* приводит лишь к занижению ассоциации.

5.1.1.2. Влияние искажающих факторов

Обнаруженную (или неподтвержденную) ассоциацию можно объяснить, допустив, что полученные результаты подвержены систематическому влиянию какого-то нераспознанного фактора. Такой неучтенный фактор именуется **искажающей переменной, конфаундером** (*confounding variable*), а ситуация именуется **искажением** (*confounding*) (Кельмансон И. А., 2002; Власов В. В., 2004; Покровский В. П., Брико Н. И., 2012; Rothman K. J. [et al.], 2008; Woodward M., 2014). Искажающий фактор должен быть ассоциирован как с патогенным воздействием, так и, независимо от него, с риском развития заболевания. Именно в такой ситуации существует вероятность, что выявленная ассоциация между воздействием и заболеванием может быть полностью или частично связана с влиянием искажающего фактора. Искажающий фактор может приводить как к недооценке, так и к переоценке выраженности реальной ассоциации.

Пример

Изучается ассоциация между образовательным уровнем матери и риском нарушений поведения ребенка первого года жизни. В качестве искажающего фактора может выступать юный возраст матери, который сам по себе влияет как на ее уровень образования, так и на риск нарушения поведения ребенка. Следовательно, выявление ассоциации между низким уровнем образования матери и риском нарушения поведения ребенка фактически может отражать преобладание среди менее образованных матерей юных женщин, и возраст матери будет искажать истинную ассоциацию между степенью ее образования и риском нарушения поведения ребенка.

Аналогично, наблюдаемая связь между потреблением матерью кофе во время беременности и риском рождения ребенка с малой массой тела может, хотя бы частично, объясняться курением матери во время беременности. Действительно, курение нередко сочетается с высоким уровнем потребления кофе, а само по себе курение также является доказанным фактором риска рождения ребенка с малой массой тела (Кельмансон И. А., 2015). Разумно предположить, что эффект употребления кофе матерью во время беременности в отношении риска рождения ребенка с малой массой тела может быть

недооценен, если не учитываются сердечно-сосудистые заболевания у матери. При сердечно-сосудистой патологии мать может реже пить кофе; сами же сердечно-сосудистые заболевания у матери являются доказанным независимым фактором риска рождения ребенка с малой массой тела. Следовательно, для матерей, употребляющих большое количество кофе во время беременности, потенциально характерна более низкая частота сердечно-сосудистых заболеваний и, как результат, возможен более низкий выявляемый риск рождения детей с малой массой тела. В случае отсутствия связи между третьим фактором и изучаемым патогенным воздействием, равно как и между этим фактором и риском заболевания, данный третий фактор уже не может выступать в качестве искажающего (рис. 54).

Уточнения относительно характеристик искажающих факторов

— Будучи прогностическим фактором высокого риска заболевания, искажающий фактор не обязательно должен иметь истинную причинную (каузальную) связь с данным заболеванием (Rothman K. J. [et al.], 2008). На практике в большинстве случаев искажающие факторы лишь коррелируют с истинными причинными факторами, обуславливающими повышенный риск того или иного заболевания (состояния). В частности, возраст и пол обследуемого практически всегда выступают в качестве наиболее вероятных искажающих факторов, однако это не означает, что в каждом случае у них этиологическая роль: чаще всего они отражают комплексное влияние истинных причин, влияющих на риск заболевания и обусловленных возрастными-половыми различиями.

— Предполагаемый искажающий фактор должен оказывать влияние на риск заболевания (патологического состояния) даже среди лиц, не подвергшихся исследуемому патогенному воздействию в данном конкретном исследовании.

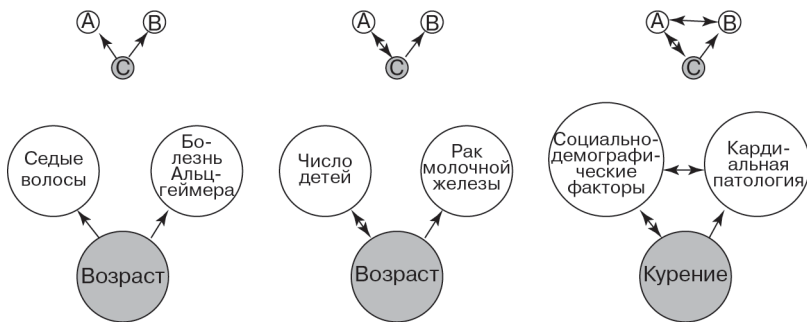


Рис. 54. Ассоциация между факторами А и В, фактор С выступает в качестве искажающего

— Предполагаемый искажающий фактор не может выступать лишь промежуточным звеном в цепи событий, связывающих патогенное воздействие и заболевание (Кельмансон И. А., 2002).

Очевидно, что в зависимости от анализируемой гипотезы один и тот же фактор может выступать в роли искажающего или не выступать в этой роли, если он включается в цепь причинно-следственных событий в качестве промежуточного звена.

Пример

Изучается влияние курения матери на риск рождения ребенка с малой массой тела. С одной стороны, известно, что курение приводит к снижению уровня гемоглобина. Поскольку анемия у матери сама по себе является одним из факторов риска малой массы тела новорожденного ребенка, в качестве этиологической гипотезы можно выстроить цепь: курение матери — анемия — малая масса тела у новорожденного. Анализ подобной гипотезы не требует рассмотрения анемии у беременной женщины в качестве искажающего фактора. В данном случае анемия может рассматриваться в качестве признака-медиатора, что обсуждалось в предыдущих разделах (Barker С. [et al.], 2002; Rothman К. J. [et al.], 2008; Woodward М., 2014). С другой стороны, предметом анализа может послужить гипотеза о независимом от анемии влиянии курения матери во время беременности на риск рождения ребенка с малой массой тела. В подобной ситуации уровень гемоглобина у беременной женщины должен учитываться в качестве возможного искажающего фактора.

На практике бывает достаточно трудно с уверенностью сказать, какие характеристики в конкретном исследовании могут выступать в качестве искажающих факторов. Далеко не в каждом случае представляется очевидной возможность ассоциации того или иного фактора с предполагаемым патогенным воздействием и риском заболевания. В подобной ситуации можно считать обоснованным соблюдение следующего методологического подхода: на основе анализа имеющихся данных исследовать нестандартизованную, грубую (*crude, unadjusted*) ассоциацию между вероятным патогенным воздействием и риском заболевания, а затем определить корректированное (*adjusted*) значение этой ассоциации с учетом влияния предполагаемого искажающего фактора. Если коррекция (*adjustment*) принципиально не влияет на полученное значение ассоциации, можно сделать вывод о том, что искажающее влияние фактора не имеет существенного значения. Если имеет место принципиальное изменение показателя ассоциации, то очевидно значимое влияние искажающего фактора на полученные результаты. Окончательное суждение об истинности обнаруженной ассоциации допустимо лишь после рассмотрения возможного влияния всех потенциальных искажающих факторов. Вот почему на стадии планирования исследования необходимо определить, какие показатели будут анализироваться в качестве потенциальных искажающих факторов, что обеспечит сбор необходимой ин-

формации, касающейся этих показателей. Иногда выполнение данной задачи требует сбора заведомо избыточной информации, так как изначально не очевидно, какие характеристики могут независимо ассоциироваться с изучаемым воздействием и риском заболевания, а следовательно, не представляется возможным заранее предусмотреть все вероятные искажающие факторы. Нужно также еще раз подчеркнуть, что выраженность выявленной ассоциации между изучаемым воздействием и риском заболевания, равно как и степень статистической достоверности этой ассоциации, еще ничего не говорит о вероятности возможного влияния искажающих факторов.

Оценивая возможное влияние искажающего фактора, важно не только констатировать его отсутствие или присутствие, но и — в случае присутствия — определить, в каком направлении он влияет на полученные результаты:

— позитивное искажение (*positive confounding*): влияние фактора приводит к тому, что вычисленная ассоциация (как положительная, так и отрицательная) оказывается более сильной, нежели реально существующая;

— негативное искажение (*negative confounding*), если оно приводит к недооценке реально существующей ассоциации.

Пример

Позитивное искажение. Возможное влияние курения матери во время беременности на ассоциацию юного возраста матери с риском нарушений поведения у детей: юные матери чаще курят во время беременности, а последнее обстоятельство сопряжено с повышенным риском нарушения поведения ребенка. Следовательно, выявляемая ассоциация юного возраста матери с риском нарушений поведения ребенка может оказаться завышенной по сравнению с реально существующей за счет искажающего влияния фактора курения матери.

Негативное искажение. Возможное влияние возраста матери на ассоциацию между отсутствием грудного вскармливания и риском нарушений сна у детей первого года жизни. С одной стороны, женщины старшего возраста имеют более высокий риск нарушений лактации и раннего перевода ребенка на искусственное вскармливание, которое ассоциировано с повышенным риском нарушений сна у ребенка. С другой стороны, риск нарушений сна ниже у детей, родившихся у женщин старшего возраста, чем у юных матерей. Следовательно, выявляемая ассоциация между искусственным вскармливанием ребенка и риском нарушений сна у ребенка может быть сглажена за счет искажающего влияния возраста матери.

Термины «негативное» и «позитивное» искажение относятся к влиянию искажающего фактора на выраженность ассоциации между изучаемым воздействием и его эффектом, а не к самой направленности ассоциации между воздействием и риском заболевания.

5.1.1.3. Минимизация влияния искажающих факторов

Можно выделить важнейшие принципы, позволяющие на стадии планирования исследования и последующей статистической обработки данных свести к минимуму влияние искажающих факторов:

- стандартизация (*standardization*);
- рандомизация (*randomization*);
- ограничение (*restriction*);
- матчинг (*matching*) (Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008; Woodward М., 2014).

Принцип рандомизации подробно рассматривался в разделе, посвященном дизайну исследования.

5.1.1.3.1. Стандартизация

Показатели распространенности того или иного заболевания или изучаемого состояния могут быть представлены применительно к обследованной популяции в целом, и в подобных случаях они являются *нестандартизованными* (*crude*). Однако возможно представление этих показателей применительно к отдельным категориям, сформированным на основе конкретных характеристик, связанных с искажающим фактором: возрастом, полом, этнической группой и т. п. В подобных случаях вычисленные значения будут показателями, *специфичными для отдельных категорий* (*category-specific rates*). Нестандартизованный показатель представляет собой суммарную характеристику, вычисляемую делением общего числа исходов в популяции (заболеваний, случаев смерти т. п.) на численность популяции в определенный период времени.

Пример

Рассмотрим данные о смертности пациентов с диагнозом болезнь Альцгеймера в США в 2000 и 2010 гг. Смертность среди пациентов всех возрастных групп составила 18,1 на 100 00 населения в 2000 г., 25,1 на 100 000 населения в 2010 г. (Alzheimer's А., 2012). Однако такие нестандартизованные показатели, свидетельствующие о нарастании смертности от данного заболевания, не дают представления о том, в какой степени указанные закономерности распространяются на людей разных возрастных групп. Известно, что для лиц старшего возраста характерен максимальный риск развития заболевания и увеличение продолжительности жизни, а следовательно, изменение возрастной структуры населения, может оказывать модифицирующее влияние на структуру смертности, в том числе на смертность от болезни Альцгеймера. Иными словами, возраст пациента может выступать в качестве искажающего фактора. В табл. 30 представлены специфичные для отдельных возрастных категорий показатели смертности, основанные на приведенных демографических данных и соответствующих расчетах. Можно

сделать вывод о том, что смертность от болезни Альцгеймера в старших возрастных группах существенно превышает общие для всех возрастов показатели смертности, а возрастание смертности от данного заболевания в 2010 г. по сравнению с 2000 г. преимущественно затрагивает возрастную группу лиц старше 75 лет.

Таблица 30

**Смертность от болезни Альцгеймера на 100 000 населения
в США в 2000 и 2010 гг.**

(Alzheimer's A., 2012; US Census Bureau, 2000, 2010 Census)

Возрастная группа, лет	2000 г.			2010 г.		
	Смертность от болезни Альцгеймера	Численность населения	Число умерших	Смертность от болезни Альцгеймера	Численность населения	Число умерших
45–54	0,2	37 677 952	75	0,3	45 006 716	135
55–64	2,0	24 274 684	485	2,1	36 482 729	766
65–74	18,7	18 390 986	2 439	19,8	21 713 429	4 299
75–84	139,6	12 361 180	17 256	184,5	13 061 122	24 098
85 +	667,7	4 239 587	28 307	987,1	5 493 433	54 226
Суммарно	50,1	96 944 389	48 562	68,6	121 757 429	83 524

Сопоставление показателей, специфичных для отдельных категорий, в том числе возрастных, во многих случаях требует анализа большого числа таких показателей, поэтому более предпочтительным является сопоставление суммарных характеристик, учитывающих различия в структуре популяций. Последнее достигается при помощи процедуры, именуемой **стандартизацией** (*standardization, adjustment*).

Стандартизованные показатели — суммарные показатели, которые учитывают различия сопоставляемых популяций по конкретным характеристикам. Выделяют прямой и непрямой методы стандартизации, сходство которых состоит в учете взвешенных средних показателей, специфичных для отдельных категорий, а различия — в методике вычисления этих весов (Кельмансон И. А., 2002; Rothman K. J. [et al.], 2008).

Прямая стандартизация (*direct standardization*). Расчеты основаны на использовании показателей, специфичных для отдельных категорий в фактически изучаемой популяции, которым придается весовое значение, исходя из структуры стандартного распределения. В качестве стандарта используют распределение наблюдений в одной

из сопоставляемых популяций, усредненное распределение наблюдений в двух сопоставляемых популяциях или некий иной стандарт (например, распределение показателей в какой-то определенный год, в какой-то определенной стране и т. п.). Очевидно, вычисляемые стандартизованные показатели будут различаться в зависимости от выбора стандарта. Однако важно, что разница между вычисленными абсолютными значениями в двух сопоставляемых популяциях будет примерно одинакова независимо от того, какой стандарт был выбран.

Пример

Сопоставляются показатели смертности от болезни Альцгеймера в США в 2000 и 2010 гг. (табл. 31). Допустим, структура популяции, наблюдавшаяся в 2000 г. (стандарт), не изменилась в 2010 г. Проведем расчет, какова была бы смертность от болезни Альцгеймера в 2010 г., если бы уровень смертности в отдельных возрастных группах соответствовал фактически наблюдаемому в 2010 г., а структура популяции была бы такой, как в 2000 г. Полученные путем прямой стандартизации показатели смертности в 2010 г. составляют 71,1 на 100 000. Это условное значение в 1,4 раза (на 40 %) превышает показатель смертности в 2000 г. для соответствующих возрастных групп. Приводимые статистические показатели для всех возрастных групп дают аналогичные результаты: в 2010 г. показатель смертности (25,1 на 100 000 населения) на 40 % превышает показатель 2000 г. (18,1 на 100 000 населения). Такие результаты позволяют говорить о том, что изменение возрастной структуры населения не оказало искажающего влияния на сопоставление показателей смертности.

Таблица 31

Сопоставление смертности от болезни Альцгеймера на 100 тыс. населения в США в 2000 и 2010 гг. по методу прямой стандартизации

Возрастной диапазон	Популяция в 2000 г.	Смертность в 2010 г.	Ожидаемое число случаев смерти в 2010 г.*
45–54	37 677 952	0,3	113
55–64	24 274 684	2,1	510
65–74	18 390 986	19,8	3 641
75–84	12 361 180	184,5	22 806
85 +	4 239 587	987,1	41 849
Суммарно	96 944 389	—	68 919

* Исходя из смертности в 2000 г.

Стандартизованный показатель: $(68\,919/96\,944\,389) \cdot 100\,000 = 71,1/100\,000$.

Непрямая стандартизация (*indirect standardization*) заключается в том, что стандартная популяция используется в качестве источника

сведений о частоте изучаемого показателя в отдельных категориях; эти показатели прилагаются к «весам» отдельных категорий в фактически изучаемой популяции. Ожидаемое число случаев заболевания или смерти в каждой фактической категории обследуемой популяции можно вычислить путем умножения специфичной для данной категории частоты показателя в стандартной популяции на «вес» каждой категории в фактической популяции. Общее число ожидаемых случаев заболевания или смерти в фактической популяции будет равно сумме ожидаемого числа случаев в каждой категории этой популяции. Полученные результаты могут быть представлены как стандартизованное отношение событий (*standardized events ratio, SER*), в частности стандартизованное отношение заболеваемости (смертности) (*standardized morbidity or mortality ratio, SMR*):

$$SMR = (\text{фактическое число случаев смерти} / \text{теоретически ожидаемое число}) \cdot 100 \%$$

Пример

Вернемся к сведениям о смертности от болезни Альцгеймера в США в 2000 и 2010 гг. (табл. 32). В качестве стандарта можно использовать показатели смертности, специфичные для отдельных возрастных категорий, наблюдавшиеся в 2000 г., и приложить их к «весам» отдельных возрастных категорий, наблюдавшихся в 2010 г. В результате получим, что в 2010 г. теоретически можно было бы ожидать 59 793 случая смерти пациентов, для сравнения: фактическое значение — 83 524. Следовательно, стандартизованное отношение смертности составит:

$$SMR = (83\,524 / 59\,793) \cdot 100 \% = 140 \%$$

Таким образом, снова приходим к выводу, что даже с коррекцией на изменение возрастной структуры населения смертность в 2010 г. выросла на 40 % по сравнению с 2000 г.

Таблица 32

Расчет показателей смертности методом непрямой стандартизации

Возрастной диапазон	Смертность в году (на 100 тыс.)		Популяция в 2010 г.	Число случаев смерти в 2010 г.	
	2000	2010		фактическое	ожидаемое*
45–54	0,2	0,3	45 006 716	135	90
55–64	2,0	2,1	36 482 729	766	730
65–74	18,7	19,8	21 713 429	4 299	4 060
75–84	139,6	184,5	13 061 122	24 098	18 233
85 +	667,7	987,1	5 493 433	54 226	36 680

* Исходя из смертности в 2000 г.

Фактическое число случаев смерти в 2010 г. = 83 524.

Ожидаемое число случаев смерти в 2010 г. = 59 793.

SMR = (83 524/59 793) · 100 % = 140 %.

Слово «стандартизованный» в проводимых расчетах означает, что была осуществлена коррекция по одному или нескольким факторам, которые могли повлиять на вычисляемые показатели. Важно, чтобы стандартная популяция была максимально приближена к фактически анализируемой в отношении иных факторов, которые также могут потенциально влиять на риск возникновения изучаемой патологии.

В зависимости от того, какого рода информацию пытается получить исследователь, принимается решение о том, какой из показателей — нестандартизованный, стандартизованный или специфичный для отдельной категории — предпочтительнее использовать.

Нестандартизованный показатель отражает истинную ситуацию в популяции. Его легко вычислить и допустимо широко использовать при сопоставлении данных, полученных в различных популяциях. В то же время нестандартизованный показатель подвержен сильному влиянию ряда искажающих факторов, связанных с различиями в структуре сопоставляемых популяций, что нередко затрудняет их корректную интерпретацию. *Стандартизованный показатель* представляет собой интегральную характеристику, нивелирующую существующие популяционные различия, что делает их весьма удобными для сопоставлений, однако такой показатель лишен реального смысла, так как вычисляются на основе произвольно выбранного стандарта.

Показатель, специфичный для отдельных категорий, не подвержен влиянию искажающих факторов и дает возможность получить наиболее полное представление о распространенности заболевания в популяции, однако проведение сопоставлений на основе указанных показателей является достаточно трудоемким из-за необходимости сравнения большого количества цифр, относящихся к отдельным анализируемым категориям.

5.1.1.3.2. Ограничение

Влияние потенциальных искажающих факторов оказывается практически исключенным, если эти факторы одинаково себя проявляют в группах, сформированных в зависимости от наличия (отсутствия) заболевания или предполагаемого патогенного воздействия. Для того чтобы достигнуть этого, доступ к участию в исследовании должен быть ограничен, участвовать могут только лица, относящиеся к той или иной категории в зависимости от их отношения к предполагаемому искажающему фактору. В этом и состоит принцип ограничения.

Пример

Если пол является предполагаемым искажающим фактором, к участию в исследовании могут быть допущены лица только определенного пола. Если

вероятным искажающим фактором является возраст, условием участия в исследовании может быть принадлежность лиц к узкому возрастному диапазону.

Принцип ограничения является достаточно простым, удобным и относительно доступным методом предотвращения или минимизации влияния искажающих факторов.

Недостатки, связанные с практической реализацией:

— Ограничение приводит к уменьшению потенциального числа наблюдений, что может сочетаться с дополнительными трудностями на пути обеспечения адекватной численности выборки и статистической силы исследования.

— Применение принципа ограничения не исключает резидуального эффекта искажающего фактора, если используемые критерии для отбора не являются достаточно строгими, а «коридор» для вхождения в исследование того или иного субъекта относительно широк. Например, в исследованиях, предполагающих возможность искажающего влияния возраста ребенка, неоправданное расширение возрастного диапазона обследуемых может оказать существенное влияние на полученные результаты.

— Реализация принципа ограничения не дает возможности оценить выраженность изучаемых ассоциаций в отдельных категориях искажающего фактора. Так, если пол является предполагаемым искажающим фактором, принцип ограничения и включение в исследование представителей какого-то одного пола может полностью нивелировать нежелательный эффект этого искажающего фактора. Однако выявленная в ходе такого исследования ассоциация не даст информации о том, в какой мере полученные закономерности распространяются на представителей противоположного пола (Кельмансон И. А., 2002).

Таким образом, увеличивая внутреннюю валидность исследования, ограничение может отрицательно сказаться на возможности обобщения полученных результатов исследования, т. е. снизить его внешнюю валидность.

5.1.1.3.3. Матчинг

Принцип матчинга, или матчирования, предполагает использование специальных приемов как на стадии планирования исследования, так и на стадии анализа данных. Если ограничение требует включать в исследование лишь тех субъектов, у которых потенциальный искажающий фактор соответствует заранее заданной фиксированной категории, то матчинг допускает участие в исследовании лиц, относящихся к различным категориям. Однако при этом каждый отдельный

обследуемый субъект должен иметь такие характеристики искажающих факторов, чтобы в целом распределение значений этих факторов было максимально сходным в сопоставляемых группах.

Пример

Проводится исследование по принципу «случай — контроль» для изучения ассоциации между курением матери во время беременности и наличием у детей нарушений сна. Во избежание возможного искажающего влияния пола и возраста, дети, относящиеся к двум сопоставляемым группам (с наличием или отсутствием нарушений сна), должны подбираться таким образом, чтобы на момент обследования распределение по полу и возрасту было одинаковым в двух группах. Тогда сопоставление частоты курения во время беременности матерей из двух групп даст значения ассоциации, не подверженной влиянию искажающего влияния пола и возраста ребенка. Практически данная задача решается на основе формирования обследуемой и контрольной группы по принципу матчинга, или копий-пар. Так, мальчику 6 лет, имеющему нарушения сна, в качестве контрольного наблюдения следует подобрать ребенка того же пола (мальчика) и возраста (6 лет), не имеющего нарушений сна. Соблюдение этого принципа подбора контрольных наблюдений для каждого обследуемого пациента из группы «случаев» приводит к идентичности распределения значений искажающих факторов в обеих сопоставляемых группах.

Несмотря на очевидные положительные стороны матчинга, связанные с устранением возможного влияния искажающих факторов, данный метод планирования исследования имеет и ряд недостатков. Прежде всего, отбор контрольных наблюдений на основе учета строго регламентированных характеристик технически сложен и может занять много времени, особенно в тех случаях, когда матчинг осуществляется с учетом многочисленных факторов. Вот почему матчинг редко используется в масштабных когортных исследованиях, где целесообразнее обследование большого числа лиц с последующим использованием специальных аналитических приемов, направленных на устранение влияния искажающих факторов, в числе которых стратификация и многофакторный анализ (эти методы изложены в последующих разделах).

Матчинг значительно чаще используется в исследованиях по принципу «случай — контроль», которые, как правило, охватывают меньшее количество наблюдений. Однако и в этих ситуациях строгое соблюдение принципов матчинга может быть достаточно затруднительным и трудоемким. Если для исследования доступно достаточное число «случаев» (лиц, имеющих изучаемое заболевание или состояние), то вполне достаточным оказывается матчинг по принципу 1 : 1, т. е. подбор одного контрольного наблюдения для каждого «случая».

Однако в тех ситуациях, когда число «случаев» оказывается ограниченным или фиксированным, увеличение статистической силы исследования может быть достигнуто за счет увеличения численности контрольной группы наблюдений, что предполагает подбор более одного контрольного наблюдения для каждого «случая» (такой принцип именуется матчингом $r : 1$). Возрастание информативности исследования на фоне увеличения числа контрольных наблюдений сохраняется лишь до тех пор, пока соотношение контрольных наблюдений и «случаев» не достигает $4 : 1$; в дальнейшем реальный эффект такого наращивания численности контрольной группы становится весьма скромным.

Не менее важной проблемой, связанной с осуществлением матчинга, является то, что его реализация не позволяет учесть влияние фактора, на основании которого осуществляется матчинг, так как распределение значений этого фактора делается искусственно одинаковым в сопоставляемых группах.

Матчинг не позволяет избежать влияния иных неучтенных искажающих факторов, за исключением тех ситуаций, когда эти факторы и факторы, по которым осуществляется матчирование, сильно коррелируют между собой.

Матчинг часто затрудняет последующее применение метода стратификации в отношении тех факторов, по которым не проводился матчинг. Несмотря на наличие статистических методов многофакторного анализа, позволяющих преодолеть указанное препятствие, к числу которых относится метод условной логистической регрессии (*conditional logistic regression*), практическое использование этих методов приводит к искусственному ограничению контингента обследуемых теми наблюдениями, которые относятся к дискордантным парам (Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008; Woodward М., 2014). Следовательно, матчинг может приводить к возникновению ряда дополнительных трудностей в ходе анализа данных.

Однако даже с учетом указанных недостатков матчинга, он остается весьма желательным в силу ряда обстоятельств. Несоблюдение этого принципа на стадии планирования исследования может привести к тому, что в сформированных группах будет затруднен подбор наблюдений, сопоставимых между собой по ряду потенциально искажающих факторов. Это, в свою очередь, может помешать реализации аналитических методов коррекции этих факторов. Например, подбор контрольных наблюдений методом матчинга по географическому распределению во многом помогает «уравнять» наблюдения с точки зрения возможного искажающего влияния комплекса экологических факторов, влияние каждого из которых в отдельности было бы

невозможно учесть и нивелировать в ходе последующего анализа в силу необходимости многоуровневой стратификации ограниченного числа наблюдений.

5.1.2. Особенности анализа данных с учетом возможного влияния искажающих факторов

5.1.2.1. Особенности анализа данных в случаях матчинга

Анализируя данные исследований «случай — контроль» с соблюдением принципа матчинга, важно понимать, какова роль матчинга в организации исследования с точки зрения его влияния на вероятные искажающие факторы. При матчинге обследуемых, относящихся к категории «случаев» и контрольных наблюдений, сопоставляемые группы становятся гомогенными в отношении ряда вероятных искажающих факторов. Если характеристики, по которым осуществляется матчинг, являются истинными искажающими факторами, процедура приведет к некоторому выравниванию выявляемых частот предполагаемых патогенных воздействий в группах обследуемых, относящихся к «случаям» и контрольным наблюдениям, чего могло бы не наблюдаться в обычной ситуации при отсутствии матчинга. Если это обстоятельство не учитывается в ходе статистического анализа, полученное значение ассоциации между патогенным воздействием и риском заболевания может оказаться заниженным.

Представлено распределение данных, полученных в результате исследования по принципу «случай — контроль» с матчингом наблюдений (табл. 33). Каждая ячейка таблицы будет отражать число пар наблюдений, относящихся к конкретной категории¹. Строки такой таблицы отражают статус «случаев» в зависимости от патогенного воздействия, а столбцы — статус контрольных наблюдений из соответствующей пары. Таким образом, ячейке *a* соответствует число пар, в которых как «случаи», так и контроль подверглись патогенному воздействию; ячейке *b* — число пар, в которых «случаи» подверглись воздействию, а контрольные наблюдения — нет; ячейке *c* — число пар, в которых контрольные наблюдения были подвержены влиянию патогенного воздействия, а «случаи» — нет; наконец, ячейке *d* — число пар, в которых ни «случаи», ни контрольные наблюдения не были подвержены воздействию.

¹ В обычной таблице вида 2×2 для данных без матчинга каждая ячейка отражает число индивидуумов в зависимости от имевшего место патогенного воздействия и наличия заболевания.

**Представление данных в четырехпольной таблице,
соответствующей исследованию по принципу «случай – контроль»
с матчингом наблюдений**

«Случай»	Контрольные наблюдения		Суммарно
	Имелось воздействие	Отсутствовало воздействие	
Имелось воздействие	<i>a</i>	<i>b</i>	<i>a + b</i>
Отсутствовало воздействие	<i>c</i>	<i>d</i>	<i>c + d</i>
Суммарно	<i>a + c</i>	<i>b + d</i>	<i>T = a+b+c+d</i>

Значения, соответствующие ячейке *a* (патогенное воздействие в группах «случаев» и контрольных наблюдений) и ячейке *d* (отсутствие патогенного воздействия в обеих группах), отражают число *конкордантных пар*, для которых статус по отношению к патогенному воздействию одинаков. Очевидно, эти конкордантные пары наблюдений не дают возможности судить о выраженности ассоциации между патогенным воздействием и риском заболевания. Такая информация может быть получена из сопоставления числа *дискордантных пар*, в которых различно отношение «случаев» и контрольных наблюдений к вероятному патогенному воздействию. Относительный риск заболевания, ассоциированный с воздействием, вычисляется как отношение числа пар, в которых «случаи» подверглись воздействию, а контрольные наблюдения — нет, к числу пар, в которых контрольные наблюдения подверглись такому воздействию, а «случаи» — нет (Hennekens С. Н. [et al.], 1987; Rothman К. J. [et al.], 2008; Woodward М., 2014):

$$RR = b/c.$$

Оценка достоверности такой ассоциации может быть осуществлена при помощи статистики χ^2 для пар, подобранных по принципу матчинга:

$$\chi^2_{(df=1)} = \frac{(b - c)^2}{b + c}.$$

Расчет ДИ значений относительного риска осуществляется по обычной методике, изложенной в главе 4.

Вычисление относительного риска может быть затруднено в тех исследованиях, когда матчинг контрольных наблюдений и случаев проводится в отношении, отличном от 1 : 1. Методика подобных расчетов изложена в специальных руководствах (Кельмансон И. А., 2002;

Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Woodward M., 2014). Следует отметить, что существенные различия, выявляемые при расчетах показателей относительного риска с учетом матчинга и без него, косвенно свидетельствуют о том, что те параметры, которые использовались для матчинга, в действительности являлись искажающими факторами. Напротив, сходство значений показателей относительного риска, рассчитанных с учетом матчинга и без него, говорит о том, что характеристики, использованные для матчинга обследуемых, в действительности не являлись искажающими факторами, что позволяет осуществлять расчеты без их учета, увеличив численность наблюдений и статистическую силу исследования. Кроме того, необоснованное, избыточное матчи́рование (*overmatching*) может приводить к снижению статистической силы исследования, особенно в исследованиях типа «случай – контроль» (Rothman K. J. [et al.], 2008; Woodward M., 2014).

5.1.2.2. Особенности стратифицированного анализа

Стратификация — это прием, используемый для учета и устранения влияния искажающих факторов. Его смысл состоит в том, что изучаемая ассоциация определяется в пределах однородных категорий (страт), характеризующих искажающий фактор. Например, если пол считать потенциальным искажающим фактором, то ассоциацию изучаемого патогенного воздействия с риском заболевания следует проводить отдельно для представителей каждого пола. В пределах каждой страты вычисленные значения ассоциации не подвержены влиянию искажающего влияния пола обследуемого, так как страта однородна (гомогенна) в отношении этого фактора. Очевидно, учет возможного влияния двух искажающих факторов требует стратификации по отношению к каждому из них, а общее число формирующихся страт будет равно числу комбинаций категорий каждого признака.

Пример

В качестве искажающих факторов рассматривается пол ребенка и его возраст (например, до 5 лет и старше 5 лет). Тогда число возможных страт будет равно 4: мальчики до 5 лет, мальчики старше 5 лет, девочки до 5 лет, девочки старше 5 лет. Оценка относительного риска в пределах каждой из указанных страт будет свободна от потенциального искажающего влияния пола и возраста ребенка. Каждая сформированная страта будет иметь вид таблицы 2×2 (см. табл. 21), и данные будут сгруппированы в виде совокупности таких таблиц, число которых равно числу страт.

Возможно вычисление относительного риска в пределах каждой сформированной страты и расчет ДИ по принципам, изложенным

в главе 4. Однако еще более интересным представляется вычисление интегрального показателя, отражающего суммарную ассоциацию, исходя из отдельных значений ассоциации в пределах каждой страты.

Пример

Использование стратифицированного анализа для исключения влияния искажающих факторов может быть проиллюстрировано данными изучения ассоциации семейного статуса матери (зарегистрированного или незарегистрированного брака) с риском нарушений сна у 95 детей 5 лет. Возраст матери на момент рождения ребенка может выступать в качестве искажающего фактора, так как вероятность незарегистрированного брака выше у юных матерей, а юный возраст матери сам по себе ассоциирован с высоким риском нарушений сна у ребенка. Для того чтобы нивелировать искажающий эффект влияния возраста матери, риск нарушений сна у ребенка, ассоциированный с незарегистрированным браком у женщин, вычислялся отдельно для каждой возрастной страты (табл. 34). Для каждой возрастной страты отмечался высокий относительный риск нарушений сна, ассоциированный с незарегистрированным браком. При этом можно было отметить неоднородность показателей относительного риска для отдельных страт: эффект негативного влияния незарегистрированного брака на риск нарушений сна у ребенка был более высоким у женщин старшей возрастной группы.

Таблица 34

Ассоциация семейного статуса женщины и риска нарушений сна у ребенка

Возраст матери на момент родов	Зарегистрированный брак	«Случаи» (нарушения сна у ребенка)	Контроль (отсутствие нарушений сна у ребенка)	Суммарно	Показатель отношения шансов нарушений сна у ребенка, ассоциированный с незарегистрированным браком
> 20 лет	Да	5	38	43	3,45
> 20 лет	Нет	10	22	32	
Всего		15	60	75	
≤ 20 лет	Да	1	6	7	1,80
≤ 20 лет	Нет	3	10	13	
Всего		4	16	20	
Суммарно для всех возрастных категорий	Да	6	44	50	2,98 (нескорректированный показатель) 3,03 (скорректированный по методу Мантела – Ханцела)
	Нет	13	32	45	
Всего		19	76	95	

Особого интереса заслуживает расчет суммарного для всех страт показателя относительного риска.

Нескорректированный (crude) суммарный показатель. Простейшим способом является расчет обобщенного показателя: все показатели будут соединены без стратификации. В примере этим показателям соответствуют цифры, приведенные в нижней части таблицы — суммарно для всех возрастных категорий. Расчет этого показателя дает значение отношения шансов, равное 2,98.

Сбалансированный, скорректированный (adjusted) показатель отражает эффект влияния стратифицирующего признака. Рассчитывается суммарный показатель относительного риска как взвешенного среднего этих показателей в отдельных стратах. Чаще всего для этого используют метод Мантела—Хансцела (Mantel—Haenszel) (Hennekens С. Н. [et al.], 1987; Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Woodward M., 2014). С использованием принятых обозначений (см. табл. 21) формула Мантела—Хансцела выглядит следующим образом:

— для расчета показателя отношения шансов:

$$OR_{MH} = \frac{\sum \frac{a \cdot d}{T}}{\sum \frac{b \cdot c}{T}};$$

— для когортных исследований:

$$RR_{MH} = \frac{\sum \frac{a \cdot (c + d)}{T}}{\sum \frac{d \cdot (a + b)}{T}}.$$

Полученная величина скорректированного показателя отношения шансов в примере составила 3,03. Она достаточно близка к нескорректированному суммарному значению отношения шансов, равному 2,98. Вообще о степени модифицирующего влияния искажающих факторов можно судить по различиям между значениями нескорректированного и скорректированного на основе стратификации показателей относительного риска.

Статистическая достоверность суммарного показателя относительного риска может быть оценена на основе критерия χ^2 (Hennekens С. Н. [et al.], 1987; Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Woodward M., 2014):

$$\chi^2_{(df=1)} = \frac{\left(\sum a - \sum \frac{(a+b) \cdot (a+c)}{T} \right)^2}{\sum \frac{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}{T^2(T-1)}}.$$

При использовании стратифицированного анализа основной интерес представляет то, как меняется характер ассоциации в зависимости от третьего (потенциально искажающего) фактора. Следовательно, любой стратифицированный анализ начинается с определения, имеется или нет модифицирующее влияние стратифицирующего признака на характер ассоциации. Когда выраженность ассоциации сильно варьирует в отдельных стратах, важнейшим компонентом стратифицированного анализа следует считать представление данных отдельно для каждой страты с расчетом относительного риска и его доверительных границ. Вычисление суммарного показателя ассоциации в данном случае не столь интересно, так как этот показатель может нивелировать колебания значений относительного риска в отдельных стратах. Тем не менее зачастую представляется целесообразным к отдельным для каждой страты значениям относительного риска дополнительно определить и суммарное значение данного показателя, свободное от влияния искажающего фактора.

В случаях негомогенности показателей в пределах отдельных страт использование приведенных выше формул неприемлемо. Используются иные принципы расчета стандартизованных показателей (Hennekens C. H. [et al.], 1987; Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Woodward M., 2014):

— для исследований «случай – контроль»:

$$RR_{std} = \frac{\sum a}{\sum \frac{b \cdot c}{d}};$$

— для когортных исследований:

$$RR_{std} = \frac{\sum a}{\sum \frac{c \cdot (a + b)}{c + d}}.$$

Признак, выступающий в качестве стратифицирующего, может быть как истинным искажающим фактором, так и простым модификатором эффекта. Он становится искажающим фактором лишь тогда, когда неодинаково распределен в сопоставляемых группах и, как следствие, оказывает влияние на вычисленные показатели ассоциации между патогенным воздействием и заболеванием. Большинство прикладных статистических программ позволяют осуществить расчеты для стратифицированного анализа (SPSS, SAS, STATA, EpiInfo, WINPEPI).

5.1.2.3. Основы многофакторного анализа

В разделе, посвященном стратифицированному анализу, показано, что он относительно легко осуществим и доступен для восприятия. Тем не менее его проведение требует категоризации наблюдений в зависимости от числа рассматриваемых факторов и их градаций. Это обстоятельство не позволяет одновременно включить в анализ большое число таких факторов или требует существенного увеличения числа наблюдений.

Пример

При изучении ассоциации нарушений сна у детей с семейным статусом матери и стратификацией по двум возрастным группам матерей число страт составит: $2 \times 2 = 4$. Оно может увеличиться, если стратифицировать возраст матери не на 2, а на большее число возрастных групп, если увеличить число категорий признака семейного статуса матери (например, зарегистрированный брак, совместное проживание без регистрации брака, мать-одиночка, разведена, вдова). Включение в анализ еще какого-то дополнительного признака в качестве потенциального искажающего фактора приведет к тому, что в ряде ячеек таблицы число наблюдений будет весьма малым или равным нулю.

Многофакторный анализ состоит в одномоментном анализе влияния нескольких переменных на результирующий показатель. Создается математическая модель, которая, по мнению исследователя, позволяет оптимально отразить связь между изучаемым воздействием и заболеванием, а также учитывает влияние вероятных искажающих и (или) модифицирующих факторов. Выбор оптимальной модели является достаточно сложной задачей, нередко предполагающей специальные знания в области математической статистики. Подробное рассмотрение этих вопросов можно найти в ряде специальных руководств (Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Howell D. C., 2013).

Многофакторный регрессионный анализ можно признать одним из доступных подходов к одновременному учету влияния нескольких факторов. В случае простой линейной регрессии связь между двумя переменными: зависимой (исходом) и независимой — имеет вид прямой линии и описывается уравнением:

$$Y = \alpha + bx,$$

где Y — зависимая переменная; α — константа уравнения регрессии (intercept); x — независимая переменная; b — коэффициент регрессии. Коэффициент регрессии равен среднему изменению значения зависимой переменной при изменении независимой переменной на

одну единицу. Выражение: $a + bx$ нередко именуется линейным предиктором.

Множественная линейная регрессия выражает линейную связь между зависимой переменной и несколькими независимыми:

$$Y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n,$$

где n – число независимых переменных; $x_1 \dots x_n$ – значения независимых переменных, вошедших в уравнение множественной линейной регрессии; $b_1 \dots b_n$ – коэффициенты множественной линейной регрессии, соответствующие отдельной независимой переменной.

Коэффициенты регрессии, входящие в линейную модель, рассчитываются на основе специальных статистических приемов таким образом, что обеспечивается максимальное приближение теоретически ожидаемых значений зависимой переменной к фактическим. Большинство современных статистических прикладных программ позволяют осуществлять такие расчеты. Значения коэффициентов линейной регрессии могут быть интерпретированы как показатели возрастания (или убывания при отрицательной величине) среднего значения зависимой переменной при увеличении значения соответствующей независимой переменной на единицу с учетом одновременного влияния прочих независимых переменных.

Пример

Рассматриваются собственные данные: симптомы постнатальной депрессии у 28 обследованных женщин, показатели их уверенности при уходе за младенцем (по данным шкалы MABS (Brazelton T. B., Nugent J. K., 2011)), возраст женщин на момент родов, пол ребенка и порядковый номер родов (табл. 35).

Таблица 35

Симптомы постанатальной депрессии по данным Эдинбургской шкалы, возраст матери на момент родов и показатели общей уверенности матери при уходе за ребенком

Номер обследования женщины	Уровень симптомов депрессии (Эдинбургская шкала)	Общая уверенность матери при уходе за ребенком (шкала MABS)	Возраст матери на момент родов	Порядковый номер родов (первородящая = 1, повторные роды = 0)	Пол ребенка (мальчик = 1, девочка = 0)
1	1	9	30	0	0
2	11	0	23	1	0
3	2	9	31	1	1
4	7	1	19	1	0
5	10	8	23	1	1

Номер обследования женщины	Уровень симптомов депрессии (Эдинбургская шкала)	Общая уверенность матери при уходе за ребенком (шкала MABS)	Возраст матери на момент родов	Порядковый номер родов (первородящая = 1, повторные роды = 0)	Пол ребенка (мальчик = 1, девочка = 0)
6	6	4	20	1	1
7	13	2	25	0	1
8	2	9	27	1	0
9	10	3	21	1	0
10	7	3	24	1	0
11	1	8	29	0	1
12	7	4	23	0	1
13	6	9	32	1	0
14	4	3	21	1	1
15	11	1	27	1	0
16	10	8	24	1	1
17	4	3	20	1	0
18	1	7	26	0	0
19	2	8	28	1	0
20	5	2	18	1	0
21	10	1	17	1	1
22	9	1	25	1	1
23	9	0	21	1	0
24	2	9	26	1	0
25	3	9	31	0	0
26	1	7	32	0	0
27	2	4	24	1	1
28	4	0	22	1	1

Применяется простой линейный регрессионный анализ, в качестве объясняющей (независимой) переменной принята уверенность женщин при уходе за ребенком, а в качестве зависимой переменной – уровень симптомов постнатальной депрессии. Анализ выявляет прямую, статистически достоверную ($P = 0,001$) ассоциацию этих показателей:

$$Y = 8,784 - 0,651 \cdot (\text{уверенность женщины при уходе за ребенком}).$$

На каждый дополнительный 1 балл роста общей уверенности женщины при уходе за ребенком в среднем приходится снижение

уровня симптомов постнатальной депрессии на 0,651 балла (рис. 55). Однако на выявленную ассоциацию предположительно могут оказывать искажающее влияние многочисленные факторы, например возраст женщины на момент родов, наличие или отсутствие предшествовавшего опыта родов (первородящая или повторнородящая), пол ребенка. Чтобы учесть возможное влияние этих факторов на показатели ассоциации между общей уверенностью женщины при уходе за ребенком и выраженностью симптомов постнатальной депрессии, можно включить возраст женщин, факт первых или повторных родов и пол ребенка, показатель уверенности при уходе за ребенком в уравнение множественной линейной регрессии в качестве объясняющих переменных. Непрерывный количественный признак «возраст матери» может быть включен в уравнение регрессии в непосредственной форме. Категориальные признаки (первородящая/повторнородящая, пол ребенка) должны быть предварительно трансформированы, и с указанной целью прибегают к двоичной кодировке признака, когда одной, произвольно выбранной градации присваивают значение 0, а альтернативной градации — 1. Вычисленное значение парциального коэффициента регрессии, соответствующего данному признаку, будет говорить о том, насколько в среднем значение зависимой переменной выше у тех индивидуумов, которые имеют градацию признака, равную 1, по сравнению с теми, у которых градация признака равна 0. В рассматриваемом примере девочкам было поставлено в соответствие значение признака, равное нулю, а мальчикам — равное единице. Повторнородящим женщинам присвоена градация, равная нулю, первородящим — равная единице. Расчет коэффициентов уравнения множественной линейной регрессии дает следующие результаты:

$$Y = 6,992 - 0,607 \cdot (\text{уверенность при уходе, балл}) + 0,995 \cdot (\text{пол}) + 0,015 \cdot (\text{возраст, лет}) + 1,042 \cdot (\text{первородящая}).$$

При коррекции по полу ребенка, возрасту матери на момент родов, порядковому роду родов выраженность симптомов депрессии уменьшается на 0,607 балла по Эдинбургской шкале с увеличением на 1 балл общей уверенности при уходе за ребенком, оцененной по шкале MABS. Парциальный коэффициент линейной регрессии, связанный с признаком «уверенность при уходе», сохраняет статистическую достоверность ($P = 0,03$). Парциальные коэффициенты регрессии, приходящиеся на другие признаки, включенные в уравнение множественной линейной регрессии, статистически недостоверны. Из сказанного можно сделать вывод: рассмотренные признаки не оказывают искажающего влияния на выявленную ассоциацию между общей уверенностью матери при уходе за ребенком и выраженностью у нее симптомов постнатальной депрессии.

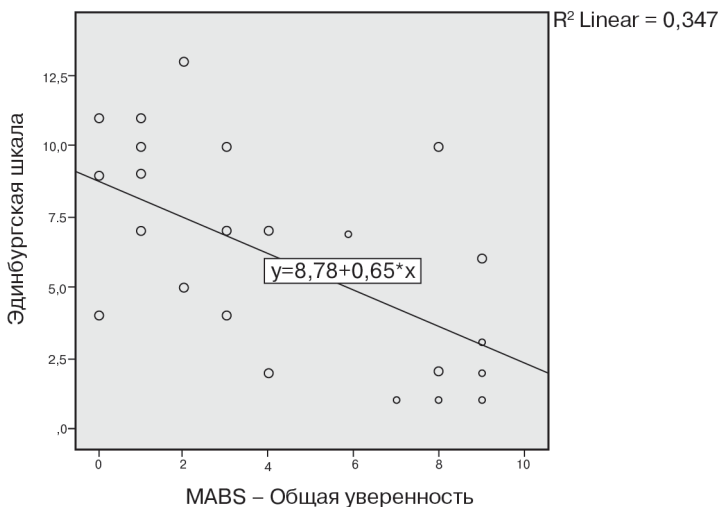


Рис. 55. Связь уровня постнатальной депрессии и общей уверенности матери при уходе за ребенком

В представленном варианте уравнения множественной линейной регрессии допускается, что по мере изменения показателей общей уверенности матери при уходе за ребенком возраст матери на момент родов, порядковый номер родов и пол ребенка сохраняют свое неизменное влияние на показатели симптоматики постнатальной депрессии. Иными словами, предполагается исключительно искажающий эффект этих факторов по отношению к общей уверенности матери при уходе за ребенком. Однако возраст матери, порядковый номер родов и пол ребенка могут по-разному проявлять себя в отношении показателей постнатальной депрессии при различных показателях уверенности матери в уходе за ребенком, т. е. выступать и в качестве модифицирующих факторов. Если учитывать подобную возможность, следует предполагать построение соответствующей модели множественной линейной регрессии. В простейшем случае в модель в качестве дополнительной переменной достаточно включить показатель взаимодействия, равный произведению взаимодействующих переменных:

$$Y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + c_1 \cdot x_1 \cdot x_2.$$

Положительное значение коэффициента регрессии, соответствующего той переменной, которая отражает взаимодействие показателей c_1 , свидетельствует об эффекте усиления, связанном с таким вза-

имодействием, а отрицательное значение — об эффекте ослабления (рис. 56). Аналогичные подходы могут быть использованы и в случае анализа вероятного взаимодействия трех и более факторов. Так, полная модель, предполагающая участие трех объясняющих переменных и все возможные варианты их взаимодействий, будет выглядеть следующим образом:

$$Y = a + b_1x_1 + b_2x_2 + b_3x_3 + c_1x_1x_2 + c_2x_1x_3 + c_3x_2x_3 + d_1x_1x_2x_3.$$

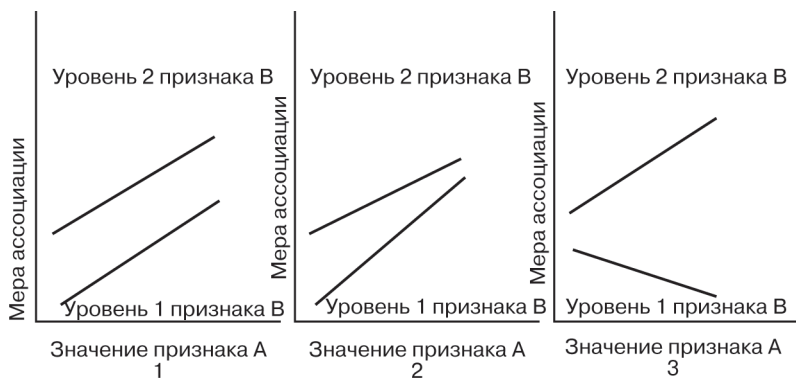


Рис. 56. Принцип взаимодействия признаков:

1 — отсутствие взаимодействия признаков А и В; 2 — отрицательное взаимодействие признаков А и В (различия ассоциации, связанные с более высоким уровнем признака В, уменьшаются по мере роста значений признака А); 3 — положительное взаимодействие признаков А и В (различия ассоциации, связанные с более высоким уровнем признака В, усиливаются по мере роста признака А)

Включение в регрессионную модель переменных, отражающих взаимодействие исходных признаков, оправдано, если это приводит к улучшению качества аппроксимации, т. е. близости рассчитываемых показателей зависимой переменной к ее реальным значениям. Кроме того, имеет смысл исследовать возможность взаимодействия признаков лишь тогда, когда выявляется статистически достоверное влияние отдельно взятых признаков на зависимую переменную. В рассматриваемом примере нет смысла изучать взаимодействие порядкового номера родов и уверенности матери при уходе за ребенком, так как порядковый номер родов сам по себе не имел статистически достоверного эффекта.

Технические трудности, связанные с осуществлением расчетов, привели к разработке компьютерных программ, в которых проводится отбор той комбинации объясняющих переменных, которая обеспечивает оптимальное качество аппроксимации. Существует несколько алгоритмов реализации подобных методов (Armitage P. [et al.], 2001; Peat J. K., Barton B., 2014).

Пошаговый регрессионный анализ (*stepwise procedure*). На первом этапе проводят испытание однофакторных линейных регрессионных моделей, учитывающих самостоятельное влияние каждой предложенной объясняющей переменной в отдельности. Выбирают ту переменную, которая обеспечивает наилучшее качество аппроксимации. Затем на втором этапе строят двухфакторные регрессионные модели путем добавления к этой переменной второй из числа тех, для которых на первом этапе анализа была выявлена статистически достоверная ассоциация. Завершается этот этап отбором комбинации, обеспечивающей наилучшую аппроксимацию. На последующих этапах по аналогичному принципу делаются попытки включения в уравнение остальных переменных до тех пор, пока наблюдается улучшение качества аппроксимации. Если дальнейшего улучшения нет, процедура пошаговой регрессии завершается.

Использование линейных регрессионных моделей возможно в тех случаях, когда зависимая переменная является количественным, непрерывным показателем, как в примере с оценкой выраженности симптомов постнатальной депрессии. Однако в исследованиях нередко возникает ситуация, когда зависимая переменная носит качественный номинальный характер и дает альтернативную оценку («болен — здоров», «выжил — погиб», «есть эффект — нет эффекта» и т. п.). Очевидно, в подобных случаях использование линейной модели для аппроксимации связи зависимой и объясняющей переменных не представляется возможным. В таких ситуациях прибегают к использованию специальной регрессионной модели.

Логистическая регрессионная модель (*logistic regression model*) (Armitage P. [et al.], 2001). Если зависимая переменная может принимать значения в диапазоне от 0 до 1, что соответствует именно тем ситуациям, когда зависимая переменная представляет собой вероятность наличия у обследуемого того или иного альтернативного качественного признака, то значения этой зависимой переменной лучше всего описываются уравнением:

$$Y = \frac{1}{1 + e^{(a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n)}},$$

где e — основание натурального логарифма; a — постоянная; $b_1 \dots b_n$ — коэффициенты регрессии; $x_1 \dots x_n$ — зависимые переменные (рис. 57). Приведенное уравнение может быть преобразовано:

$$\ln\left(\frac{Y}{1-Y}\right) = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_n \cdot x_n.$$

Правая часть полученного уравнения является традиционным линейным предиктором, характерным для уравнения множественной линейной регрессии. Левая часть представляет собой преобразованную форму зависимой переменной — натуральный логарифм отношения вероятности наличия распознаваемого состояния (например, заболевания) к вероятности его отсутствия, т. е. логарифм отношения шансов (логит). Это и есть уравнение логистической регрессии, которое отражает линейную зависимость между логитом распознаваемого состояния и совокупностью предложенных независимых объясняющих переменных. Решение уравнения логистической регрессии требует выполнения сложных математических расчетов, и большинство современных прикладных статистических программ (SAS, STATA, SPSS, STATISTICA, EpiInfo) позволяют осуществить такие расчеты.

Интерпретация коэффициентов уравнения логистической регрессии достаточно проста. По определению, эти коэффициенты соответствуют той величине, на которую увеличивается или уменьшается значение логарифма отношения шансов при увеличении значения соответствующей объясняющей переменной на единицу. Следовательно, каждый коэффициент уравнения показывает, в какой мере соответствующий ему показатель влияет на логарифм отношения шансов изучаемого зависимого признака с учетом эффекта одновременного влияния прочих объясняющих переменных, включенных в уравнение регрессии. Эти коэффициенты могут быть трансформированы в показатели отношения шансов, которые следует рассматривать как скорректированные показатели относительного риска, учитывающие влияние потенциальных искажающих факторов.

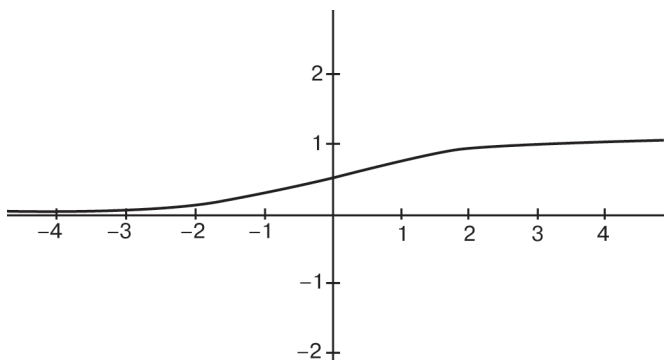


Рис. 57. Логистическая регрессия

Если независимая переменная, вошедшая в уравнение множественной логистической регрессии, имеет двоичный характер (например, пол обследуемого, наличие или отсутствие какого-то симптома), то она нуждается в предварительной трансформации (0 – отсутствие признака, 1 – наличие признака). Соответствующий этой фиктивной (dummy) переменной коэффициент логистической регрессии будет отражать скорректированный показатель относительного риска, ассоциированного с наличием изучаемого признака (Armitage P. [et al.], 2001; Rothman K. J. [et al.], 2008; Peat J. K., Barton B., 2014; Woodward M., 2014):

$$OR(x_1) = e^{b_i}.$$

ДИ вокруг этого показателя относительного риска может быть рассчитан, исходя из значения стандартной ошибки коэффициента регрессии:

$$95\% CI = e^{(b_i \pm 1.96 \cdot SEb_i)}.$$

Логистическая регрессия позволяет учитывать взаимодействие объясняющих переменных.

Практический интерес представляет осуществление многофакторного анализа с целью определить, насколько выявленная ассоциация между двумя переменными подвержена влиянию вероятных переменных – медиаторов и модераторов. Возможная ассоциация между двумя переменными с участием переменной-медиатора представлена на рис. 58. Принципиальная ассоциация, которая является предметом анализа, обозначена как путь y_3 , связывающий независимую и зависимую переменные. Путь, проходящий через переменную-медиатор, состоит из двух отрезков: y_1 – отрезок, который связывает независимую переменную и переменную-медиатор; y_2 – отрезок, который связывает переменную-медиатор и зависимую переменную. Для обоснования того, что переменная может выступать в качестве медиатора, необходимо:

- убедиться в наличии статистически достоверной ассоциации независимой переменной и переменной-медиатора (Baron R. M., Kenny D. A., 1986);

- выявить достоверную ассоциацию между предполагаемым медиатором и зависимой переменной (предполагается наличие достоверной ассоциации между независимой и зависимой переменными (путь y_3));

- убедиться в том, что при одновременном использовании путей y_1 и y_2 качество предсказания значений зависимой переменной будет лучше, чем при использовании пути y_3 . С математической точки зрения это означает сопоставление качества аппроксимации двух моделей:

первая модель:

$$\begin{cases} y_1 = a_1 + b_1 \cdot x_1, \\ y_2 = a_2 + b_2 \cdot x_2, \end{cases}$$

где первое и второе уравнения являются математическим выражением путей y_1 и y_2 соответственно;

вторая модель:

$$y_3 = a_3 + b_3 \cdot x_3,$$

где представлено математическое выражение пути y_3 .

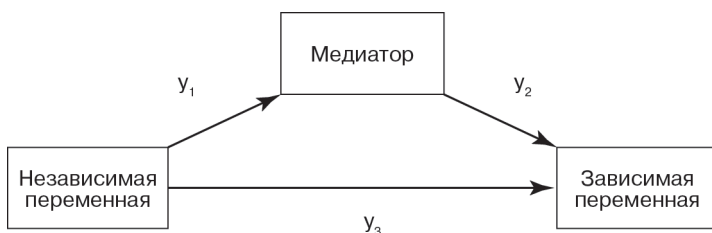


Рис. 58. Схематическое представление возможного участия переменной-медиатора в выявленной ассоциации

Методика соответствующих расчетов подробно изложена в специальных руководствах (Hayes A. F., 2013; Howell D. C., 2013). На подобные расчеты ориентированы некоторые прикладные статистические программы, например специально разработанный и доступный модуль PROCESS для программы SPSS (www.hayes.com).

Пример

Рассмотрим собственные данные о связи нарушений сна у 34 детей школьного возраста с уровнем их общих адаптивных навыков. Выраженность нарушений сна исследовалась при помощи опросника для родителей CSHQ, уровень адаптивных навыков ребенка — при помощи опросника для учителей DABS. Одновременно учитывались нарушения поведения ребенка на основании опроса учителей с использованием опросника Ахенбаха (CBCL) (табл. 36). Выявлена достоверная связь ($P < 0,001$) уровня нарушений сна и степени развития адаптивных навыков, которая описывается уравнением регрессии.

**Выраженность нарушений сна, нарушений поведения
и адаптивных навыков у детей**

Номер обследованного ребенка	Уровень общих нарушений сна у ребенка	Уровень общих адаптивных навыков ребенка	Уровень общих нарушений поведения ребенка
1	37	225	1
2	45	208	12
3	38	219	0
4	44	218	21
5	38	216	1
6	50	211	6
7	35	223	2
8	50	205	13
9	49	222	2
10	37	223	7
11	46	215	0
12	35	207	7
13	52	204	13
14	46	219	1
15	42	224	0
16	38	205	7
17	37	220	0
18	38	222	0
19	56	211	14
20	40	225	1
21	37	217	0
22	37	218	1
23	41	217	0
24	58	182	3
25	43	224	0
26	39	218	1
27	37	225	1
28	45	217	6

Номер обследованного ребенка	Уровень общих нарушений сна у ребенка	Уровень общих адаптивных навыков ребенка	Уровень общих нарушений поведения ребенка
29	45	210	2
30	44	212	1
31	33	220	2
32	36	214	1
33	39	219	0
34	42	218	4

Уровень развития адаптивных навыков = $250,8 - 0,84$ (уровень нарушений сна).

Уровень развития адаптивных навыков снижается по мере нарастания нарушений сна у ребенка. В то же время известно, что нарушения сна у детей могут негативно сказываться на выраженности нарушений поведения в дневное время, что было подтверждено и данными регрессионного анализа в настоящем исследовании ($P = 0,008$).

Уровень общих нарушений поведения = $-11,9 + 0,38 \times$ (выраженность нарушений сна).

Кроме того, выявлена статистически достоверная ($P = 0,032$) отрицательная ассоциация между выраженностью нарушений поведения ребенка и степенью развития адаптивных навыков.

Уровень развития адаптивных навыков = $217,9 - 0,61 \times$ (выраженность нарушений поведения).

Следовательно, выраженность нарушений поведения ребенка может выступать в качестве переменной-медиатора, опосредуя влияние нарушений сна на выраженность адаптивных навыков у детей. Для оценки такой возможности проведен многофакторный анализ при помощи модуля PROCESS с использованием программы SPSS (рис. 59). Результаты свидетельствуют о том, что непрямой эффект (т. е. с участием переменной-медиатора) был статистически недостоверен. Так, в многофакторной модели, где нарушение сна (TotCSHQ) и нарушение поведения (TotCBCL) выступали в качестве независимых переменных, а уровень адаптивных навыков (Adapt) — в качестве зависимой, парциальный коэффициент регрессии, приходящийся на признак TotCBCL, был статистически недостоверен ($P = 0,435$). Прямое влияние признака TotCSHQ на признак Adapt было статистически достоверно ($P = 0,017$). Таким образом, в рассматриваемом примере эффект медиатора не был существенным.

Run MATRIX procedure:

***** PROCESS Procedure for SPSS Release 2.13.1 *****

Written by Andrew F. Hayes, Ph.D. www.afhayes.com
Documentation available in Hayes (2013). www.guilford.com/p/hayes3

Model = 4
Y = Adapt
X = TotCSHQ
M = TotCBCL

Sample size
34

Outcome: TotCBCL

Model Summary	R	R-sq	MSE	F	df1	df2	p
	,4457	,1987	22,3107	7,9333	1,0000	32,0000	,0082

Model	coeff	se	t	p	LLCI	ULCI
constant	-11,9985	5,6755	-2,1141	,0424	-23,5593	-,4377
TotCSHQ	,3765	,1337	2,8166	,0082	,1042	,6487

Outcome: Adapt

Model Summary	R	R-sq	MSE	F	df1	df2	p
	,6115	,3739	48,6627	9,2566	2,0000	31,0000	,0007

Model	coeff	se	t	p	LLCI	ULCI
constant	248,2862	8,9482	27,7472	,0000	230,0359	266,5365
TotCBCL	-,2065	,2611	-,7909	,4350	-,7390	,3260
TotCSHQ	-,7571	,2205	-3,4335	,0017	-1,2068	-,3074

***** DIRECT AND INDIRECT EFFECTS *****

Direct effect of X on Y						
Effect	SE	t	p	LLCI	ULCI	
	-,7571	,2205	-3,4335	,0017	-1,2068	-,3074

Indirect effect of X on Y				
Effect	Boot SE	BootLLCI	BootULCI	
TotCBCL	-,0777	,1655	-,4598	,1609

***** ANALYSIS NOTES AND WARNINGS *****

Number of bootstrap samples for bias corrected bootstrap confidence intervals:
1000

Level of confidence for all confidence intervals in output:
95,00

----- END MATRIX -----

Рис. 59. Распечатка результатов процедуры PROCESS, выполненной при помощи программы SPSS. Adapt — показатель уровня адаптивных навыков, TotCBCL — общая выраженность нарушений поведения ребенка, TotCSHQ — общая выраженность нарушений сна у ребенка

Предметом критического анализа и рассмотрения полученных результатов является внешняя валидность исследования и угрозы этой валидности. Оценка достоинств и недостатков исследования предполагает постановку вопроса: в какой мере полученные результаты могут быть распространены за пределы самого исследования? Это и составляет суть внешней валидности исследования. Изучение внешней валидности предполагает исследование репрезентативности исследования, т. е. определение границ, касающихся людей, условий и времени, в пределах которых можно применить результаты исследования. Любая специфика обследованной выборки пациентов, процедур или времени проведения исследования снижает внешнюю валидность. Дилемма любого исследования заключается в том, что потребность увеличить внешнюю валидность может противоречить необходимости повысить валидность статистических выводов. Например, один из подходов к снижению систематической ошибки исследования и повышению валидности статистических выводов — это формирование обследуемой выборки из гомогенной целевой популяции. Однако это сделает выборку менее репрезентативной и приведет к снижению внешней валидности исследования.

Принцип репликации (воспроизведения). Чем больше степень повторяемости исходных результатов в различных условиях, тем более убедительными выглядят эти результаты. Принцип представляет собой оптимальный подход к обоснованию внешней валидности исследования. Можно выделить три вида репликации:

— *буквальная репликация (literal replication)*, которая является точным повторением исследования, осуществленным тем же исследователем с использованием идентичных процедур;

— *операционная репликация (operational replication)*, осуществляемая другими исследователями с использованием тех же методов, которые описаны исследователями в оригинальной работе;

— *репликация конструкта (constructive replication)*, которая воспроизводит основную идею исследования, но использует при этом другие методы (другую популяцию, альтернативные подходы к измерению того же конструкта) (Barker С. [et al.], 2002). Если результаты первоначального исследования принципиально повторяются в ходе нескольких таких воспроизведений, можно судить, на какие ситуации и каких лиц можно распространить полученные результаты.

Угрозы для внешней валидности исследования, которые требуют критической оценки:

— *Взаимодействие каузальной связи с единицами измерения:* эффект, выявленный при использовании одних единиц измерения, может не определяться при использовании других единиц.

— *Взаимодействие каузальной связи с вариациями легения*: эффект, выявленный в отношении одной вариации лечения, может не подтверждаться в отношении других вариаций лечения, а также в тех случаях, когда это лечение сочетается с другими формами лечения или если оно применяется лишь частично.

— *Взаимодействие каузальной связи с исходом*: эффект, выявленный в отношении одного наблюдаемого исхода, может не выявляться при фиксации другого исхода.

— *Взаимодействие каузальной связи с условиями*: эффект, выявленный в одних условиях, может не определяться при других условиях.

— *Переменная-медиатор, зависящая от контекста*: переменная-медиатор, обеспечивающая эффект в одном контексте, может не выполнять свою роль в другом (Campbell D. T., Stanley J. C., 1963; Shadish W. R. [et al.], 2001).

Исследовательские программы часто начинаются с лабораторных исследований, характеризующихся низкой внешней валидностью. Если такие начальные лабораторные результаты оказываются успешными, исследователю следует осуществить экологически более валидные исследования, позволяющие повысить степень доверия результатам.

5.2. Принципы мета-анализа

На научные исследования в области клинической психологии распространяется так называемая вероятностная парадигма. Возникновение того или иного состояния или исхода у каждого конкретного испытуемого полагается определять с той или иной вероятностью. Следовательно, клинико-психологические исследования должны быть основаны на оценке статистической достоверности выявленного эффекта. Зачастую психологическое сообщество принимает тот или иной научный факт и вырабатывает конкретные рекомендации на основе литературных обзоров, подготовленных экспертами, и принципов консенсуса, достигаемого признанными специалистами в конкретной области. Однако следует признать, что такой подход больше относится к области научной психологии, а не научной методологии. Результаты исследований не могут рассматриваться как строгий научный факт, если не будет обеспечена возможность воспроизвести их в ходе повторных испытаний в условиях, максимально приближенных к исходным (Taylor C. B., 2010). Когда речь идет о клинических или клинико-психологических проблемах, наличие определенного числа отдельных публикаций в той или иной области должно предполагать возможность некоего синтеза этих на-

копленных разрозненных данных, а значит, возникает потребность в разработке методики подобного синтеза. Подходы к оценке достоверности гипотезы отдельно взятого исследования (*hypothesis-testing model*) были детально разработаны классиками математической статистики (несколько подходов рассмотрены в предыдущих главах). Этого нельзя сказать в отношении подходов к синтезу, объединению результатов многочисленных отдельных исследований. Указанное обстоятельство послужило основанием для разработки определенных методологических принципов, направленных на объединение накопленных данных отдельных исследований. Подобный подход получил название «мета-анализ» (*meta-analysis*).

Мета-анализ — специфический методологический подход с целью объединить результаты ранее выполненных исследований (Власов В. В., 2001; Кельмансон И. А., 2011; Kulinskaya E. [et al.], 2008; Littell J. H. [et al.], 2008). Подобный подход приводит к существенному увеличению статистической силы, повышению точности оценок выраженности анализируемого эффекта (как положительного, так и отрицательного), а в случаях кажущейся противоречивости результатов отдельных исследований позволяет получить именно ту обобщенную оценку, которая вытекает из результатов всех проанализированных исследований и наилучшим образом отражает реальное положение вещей (D'Agostino R. B., Weintraub M., 1995). Сам термин «мета-анализ» был впервые использован в 1976 г. (Glass G. V., 1976). Поначалу методика мета-анализа получила применение в социологических и педагогических науках, однако впоследствии была с успехом использована в медицинских и медико-психологических исследованиях. На сегодняшний день мета-анализ используется для того, чтобы эффективно внедрять в практическую деятельность результаты научных достижений, придерживаясь принципов доказательности выводов.

Мета-анализ призван обеспечить синтез информации, доступной в данный момент времени и в связи с той или иной конкретной проблемой. Его проводят обязательно максимально объективно, независимо от того, кто его проводит, и на основе научно обоснованной и воспроизводимой методики. Следовательно, методы мета-анализа должны быть верифицируемыми и воспроизводимыми, итогом их применения должно быть получение обобщающего показателя выраженности изучаемого эффекта. Для того чтобы удовлетворять всем поставленным задачам, мета-анализ должен:

- базироваться на всей доступной информации, что требует масштабного изучения исследований в профильной области;
- последовательно и строго реализовывать выработанный подход в соответствии с заранее разработанным протоколом;
- адекватно использовать аппарат статистического анализа.

Мета-анализ основан на строгом и аргументированном отборе анализируемых исследований и обеспечении возможности воспроизвести результаты такого отбора. Используемые в ходе мета-анализа статистические подходы позволяют учесть вероятностный характер результатов, полученных как в отдельном исследовании, так и при обобщении данных всех исследований, а также количественно оценить эффект изучаемого воздействия.

Принятая классификация предполагает выделение важнейших вариантов проведения мета-анализа:

- мета-анализ литературных данных;
- мета-анализ, основанный на всестороннем изучении данных реферативных сообщений;
- мета-анализ, основанный на изучении данных индивидуальных пациентов (Кельмансон И. А., 2011).

Важнейшие задачи мета-анализа

- Повышение статистической силы за счет увеличения числа наблюдений;
- улучшение точности оценки выраженности изучаемого эффекта;
- устранение сомнений в случаях выявления противоречивых результатов исследований;
- глобальная оценка и повышение возможности обобщения результатов;
- объяснение вариабельности результатов, наблюдаемых в различных исследованиях (в частности, вследствие присутствия систематической ошибки);
- осуществление анализа в подгруппах;
- проведение исследования с учетом предполагаемых возможных противоречий с результатами других исследований в изучаемой области;
- констатация недостаточной полноты данных в изучаемой области и организация новых исследований;
- ответы на вопросы, которые исходно не были затронуты в анализируемых исследованиях.

Мета-анализ наиболее оправдан в следующих ситуациях:

- отдельные исследования по рассматриваемой проблеме основаны на малой выборке, что не обеспечивает надежности результатов и обоснованности выводов о наличии или отсутствии изучаемого эффекта;
- проведение исследования с участием большого числа пациентов невозможно или затруднительно;
- результаты исследований в указанной области противоречивы или малоубедительны (Fagard R. H. [et al.], 1996).

Практическая реализация мета-анализа требует последовательного прохождения ряда важнейших этапов:

Первый этап. Необходимо четко сформулировать ту проблему, которая должна стать предметом мета-анализа. От формулировки проблемы во многом будет зависеть и дизайн мета-аналитического исследования.

Второй этап. Следует найти доступные публикации по теме и провести их критический отбор. Задачей является изучение тех источников, в которых может содержаться исходная информация, необходимая для мета-анализа.

Третий этап. В отобранных публикациях нужно найти те статистические данные, которые необходимы для последующей обобщающей оценки выраженности изучаемого эффекта.

Четвертый этап. Проводятся статистические расчеты. Данная тема подробно рассматривается в специальных руководствах (Кельмансон И. А., 2011; Borenstein M. [et al.], 2009). Расчеты для мета-анализа позволяют осуществить некоторые прикладные статистические программы:

— доступная программа RevMan, разработанная Кокрановским сотрудничеством и предназначенная для подготовки систематических обзоров и реализации мета-анализа (<http://tech.cochrane.org/revman/download>) (Higgins J. P. T. [et al.], 2008);

— доступная программа MIX позволяет осуществить серию расчетов и построить графики, необходимые для реализации мета-анализа в среде Excel (<http://www.meta-analysis-made-easy.com>) (Bax L. [et al.], 2006);

— коммерческие программы, предназначенные для мета-аналитических расчетов, в частности Meta-DiSc (Zamora J. [et al.], 2006), Comprehensive Meta-Analysis (CMA) (Borenstein M. [et al.], 2005).

В целом нужно отметить, что принципы статистических расчетов, применяемых в мета-анализе, во многом аналогичны тем, что используются в методе стратификации. Каждая отдельная публикация выступает в качестве страты с соответствующей выраженностью эффекта, что позволяет оценить гомогенность результатов отдельных исследований и вычислить обобщающий показатель выраженности эффекта.

Пятый этап. Осуществляется подготовка отчета и его публикация — систематический обзор (*systematic review*). Отметим, что в случае проведения мета-анализа и подготовки систематического обзора можно с полным основанием рассматривать анализ литературы как самостоятельную задачу исследования и делать вывод, основанный на выполнении этой задачи.

Для успешной реализации всех указанных этапов требуются многочисленные ресурсы. Прежде всего, необходимы средства, обеспечивающие возможность библиографического поиска. Желателен

постоянный доступ к электронной библиографической базе данных, так как в ходе работы с литературой могут быть обнаружены публикации, относящиеся к предмету мета-анализа, имена авторов, названия исследований в указанной области. Для оценки качества публикаций и их критического отбора целесообразно привлечь еще одного исследователя («второго читателя»). При этом нужно помнить, что на качественное критическое прочтение одной статьи требуется от 1 до 3 ч. Нужно выделить время для совместного обсуждения мнений о качестве отбираемых работ. Необходимо предусмотреть переписку с другими исследователями с целью найти интересующие работы, верифицировать полученные сведения и дополнить недостающую информацию. Наконец, большой объем работы сопряжен с проведением статистического анализа.

Вопросы для самоконтроля

1. В чем принципиальные различия ассоциации и причинно-следственной связи?
2. Что такое внутренняя валидность исследования и каковы важнейшие угрозы внутренней валидности?
3. Что такое систематическая ошибка? Каковы ее виды?
4. Что такое искажающие факторы? Как они могут повлиять на результаты исследования? Каковы подходы к минимизации их влияния?
5. Каковы подходы к анализу данных с учетом влияния искажающих факторов?
6. Что такое внешняя валидность исследования? Каковы важнейшие угрозы внешней валидности?
7. Что такое мета-анализ? Каковы его принципы?

ПИСЬМЕННОЕ ПРЕДСТАВЛЕНИЕ СВЕДЕНИЙ ОБ ИССЛЕДОВАНИИ

Научное исследование, как правило, ориентировано на широкую профессиональную аудиторию. Обычно итог исследования представляется на суд коллег. Традиционно это предполагает письменное представление сведений о проведенном исследовании в виде журнальной статьи, научного отчета, дипломной работы, диссертационного исследования. Несмотря на многообразие указанных форм, на сегодняшний день выработана некая общая схема такого письменного представления сведений, имеющая ряд нюансов в отдельных случаях.

Обычная последовательность разделов, применяемая большинством авторов, такова:

- обзор литературы (введение для статей);
- материал и методы исследования;
- результаты исследования;
- обсуждение результатов;
- заключение;
- выводы;
- практические рекомендации;
- приложения (если есть);
- список литературы (Марьянович А. Т., Князькин И. В., 2009;

Peat J. K., 2002).

6.1. Обзор литературы

Подготовка обзора литературы по изучаемой проблеме была рассмотрена в главе 1. Знакомство с литературой по теме исследования чаще всего отражено в обзоре литературы. Такой термин применяется в дипломных работах, диссертационных исследованиях. В журнальных статьях ему чаще соответствует более компактное введение (Introduction). Введение призвано дать ответы на следующие вопросы:

- Что мы знаем?
- Чего мы не знаем?
- Зачем было проведено это исследование? (Peat J. K., 2002).

В тезисах конференций, опубликованных на английском, соответствующий раздел чаще всего обозначается термином Background (т. е. «фон»), в нем дается буквально несколько предложений, характеризующих суть проблемы.

Задача литературного обзора, представленного в качестве первой главы дипломной работы или диссертационного исследования, — не анализ всей литературы, относящейся к данной проблеме, но обоснование необходимости проведенного исследования (Марьянович А. Т., Князькин И. В., 2009). Главное в обзоре литературы — его внутренняя логика, подводящая читателя к мысли о необходимости исследования, за которое берется автор. В первой главе желательно дать название чуть шире названия самой работы и отразить современное состояние проблемы. Саму главу можно разбить на подразделы и дать им соответствующие подзаголовки. Главное правило написания литературного обзора — «...основные идеи и их развитие» (Максименков А. Н., 1972). После изложения любого сколько-нибудь крупного вопроса, особенно если существуют противоположные точки зрения, исследователю надо высказать свое обоснованное мнение. По ходу литературного обзора исследователю нужно ссылаться на авторов соответствующих публикаций. Ссылки оформляются в соответствии с принятыми правилами цитирования, которые могут различаться в разных изданиях. При написании дипломной и диссертационной работ авторам следует ориентироваться на действующий ГОСТ Р7.0.5-2008 (<http://protect.gost.ru/v.aspx?control=7&id=173511>). Напомним также, что ряд библиографических прикладных компьютерных программ позволяет осуществлять форматирование ссылок и создавать список литературы в соответствии с указанными стандартами (в частности, программа Mendeley). В конце главы «Обзор литературы» желательно поместить резюме размером не более одной страницы. В нем нужно указать, к каким умозаключениям пришел исследователь в ходе ознакомления с работами других авторов: что известно, что можно предположить с высокой степенью вероятности и как это предположение проверить исследовательскими методами. Таким образом, формулируется главная гипотеза, проверке которой будет посвящено исследование, а также задачи исследования.

6.2. Материал и методы исследования

Начать надо с описания объекта исследования. Нужно представить важнейшие социально-демографические и клинические характеристики обследованных лиц, критерии включения и исключения в исследование (пол, возраст, массу тела, диагноз, продолжительность заболевания, предшествующее лечение и т. п.). Необходимо

пояснить, чем продиктован выбор именно этого объекта и те или иные критерии включения и исключения, где и в каких условиях проводилось исследование. Нужно описать и обосновать дизайн исследования. Требуется обоснование численности обследованной выборки и подходов к ее расчету. При описании обследованных лиц надо избегать какой-либо конфиденциальной информации, которая позволила бы идентифицировать этих субъектов. Далее необходимо привести полный перечень методик, использованных в работе. Любые измерения, приведенные в разделе «Результаты исследования», должны быть получены при помощи методик, описанных в разделе «Материал и методы» (Peat J. K., 2002). На первые места ставят те методики, которые применялись ко всем обследованным лицам, а потом и те, что охватили ограниченный круг испытуемых. Если речь идет о стандартных методиках, подробно описанных в литературе, достаточно назвать и дать ссылки на соответствующие литературные источники. Все прочие методики, особенно авторские или модифицированные автором, следует описать более подробно. Говоря о собственных модификациях методов исследования, уместно привести номера патентов, авторских свидетельств или свидетельств о принятых рационализаторских предложениях. Нужно спросить себя, сможет ли читатель, ранее не знакомый ни с одним из примененных в работе методов, при желании точно воспроизвести исследование, чтобы убедиться в достоверности представленных результатов. Если нет, следует переделать текст, пока изложение не станет полным и абсолютно ясным. При наличии сомнений в том, что описываемая методика известна большинству читателей, ее надо описывать как малоизвестную (Марьянович А. Т., Князькин И. В., 2009). Иногда исследователь чувствует необходимость объяснить, почему именно эту, а не иную методику он выбрал для своей работы. Если для объяснения достаточно нескольких строк, такой фрагмент может быть включен в раздел «Материал и методы». Если же выбор метода требует сопоставления литературных данных и более пространственных рассуждений, есть основания перенести такой анализ в «Обзор литературы», куда также можно включить небольшой фрагмент с примерным названием «Методы, применяемые в исследованиях (такого-то феномена)». Последний абзац этого фрагмента должен содержать заключение автора о том, который из методов представляется ему наиболее подходящим для решения задач его исследования. Изложение теоретических основ примененных методов должно быть таким, чтобы убедить читателя в адекватности выбранных методик.

В дипломных и диссертационных работах часто приводится таблица с названием «Объем и методы исследования» или «Общая характеристика исследования». В ней дан перечень всех использо-

ванных методик и количество лиц, обследованных при помощи соответствующей методики, количество произведенных измерений, проб и т. п.

Необходимо указать, какие статистические методы обработки данных использовались в исследовании, какие прикладные статистические программы для этого применялись, каковы были критерии констатации статистической достоверности выявленных различий и ассоциаций.

По принятым правилам, раздел «Материал и методы» должен также содержать информацию о соответствии проведенного исследования принципам биоэтики, наличии информированного согласия обследованных лиц на участие в исследовании, а также сведения о наличии или отсутствии конфликта интересов.

6.3. Результаты

Раздел «Результаты исследования» является наиболее важным, так как его функция — дать ответы на вопросы, которые стояли перед исследованием. Примерная структура изложения результатов представлена на рис. 60. Оптимальным вариантом изложения результатов можно считать последовательное продвижение от описательной (однофакторной) к аналитической статистике (двухфакторному и многофакторному анализу). Такое изложение должно соответствовать логике исследования и наглядно демонстрировать ее. Размер раздела зависит исключительно от того, сколько результатов может представить исследователь, а не от того, как много он хочет о них рассказать (Peat J. K., 2002).

Первая часть раздела должна содержать исчерпывающую информацию о характеристиках обследованных субъектов, что позволит судить о возможности обобщения результатов: читателя будет интересовать, насколько обследованные лица схожи с теми, с которыми ему самому приходится сталкиваться, и в отношении какого контингента лиц применимы полученные результаты.

Вторая часть непосредственно затрагивает те вопросы, на которые призвано ответить исследование. В ней излагаются преимущественно результаты проверки гипотезы, сформулированной во введении (обзоре литературы). Следует приводить лишь те сведения, которые должны узнать читатели. Не нужно поддаваться соблазну делать какие-то пространные комментарии к приводимым фактам или включать в этот раздел какие-то аналитические процедуры, не относящиеся к основной задаче. Не следует повторять в тексте цифры, представленные в таблицах и на рисунках, а информация, содержащаяся в рисунках и таблицах, не должна дублироваться. Можно

привлечь внимание читателя, подкрепляя изложение того, что было выявлено, ссылкой на соответствующую таблицу или рисунок, подтверждающие высказанные положения. Например, «...как явствует из таблицы, дети, страдающие бронхиальной астмой, характеризовались статистически достоверно более низкими показателями качества жизни». При этом в таблице будут приведены конкретные цифры, отражающие показатели качества жизни у детей с диагнозом «бронхиальная астма» и у клинически здоровых детей из контрольной группы наблюдений, а также соответствующее значение P , отражающее уровень статистической значимости выявленных различий. Таким образом, читателю будет проще разобраться в информации, содержащейся в таблице или на рисунке, и он не остается один на один с цифрами и графиками, которые ему придется интерпретировать самостоятельно.

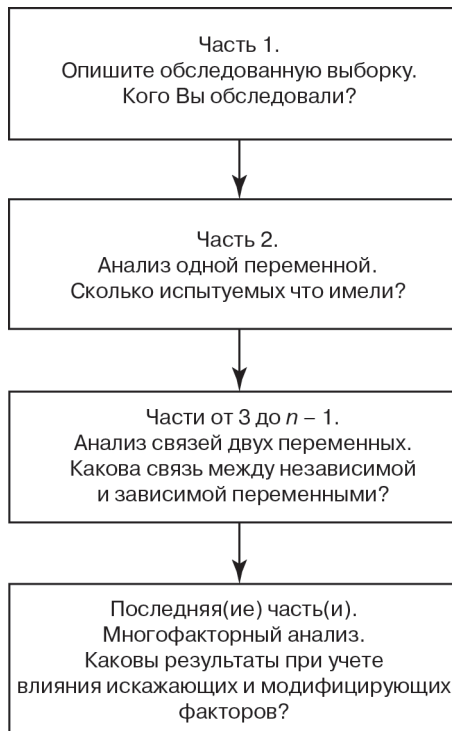


Рис. 60. Логическая структура построения главы «Результаты исследования» (Peat J. K., 2002)

Если возникает потребность сопоставить собственные результаты с результатами других исследователей, то это лучше сделать в следующем разделе «Обсуждении результатов».

Традиционно результаты и их обсуждение обычно излагаются в разных разделах (главах). Если же представляемый материал настолько обилён, что невозможно изложить его в одной главе, а обсудить — в другой, или если материал отчетливо делится на несколько смысловых частей, можно выбрать монографический способ изложения материала (Марьянович А. Т., Князькин И. В., 2009). Тогда результаты и обсуждения будут располагаться в нескольких главах, например:

Глава 3. Результаты исследования XXX.

Глава 4. Обсуждение результатов исследования XXX.

Глава 5. Результаты исследования YYY.

Глава 6. Обсуждение результатов исследования YYY.

Другой вариант монографической структуры:

Глава 3. Результаты и обсуждение исследований XXX.

Глава 4. Результаты и обсуждение исследований YYY.

Глава 5. Результаты и обсуждение исследований ZZZ.

По возможности предпочтение следует отдавать традиционной схеме изложения.

6.4. Обсуждение результатов

Анализ полученных данных, отраженный в «Результатах исследования, дает базисную информацию. Однако эти результаты нуждаются в интерпретации и обсуждении, которые позволяют понять их смысл и дать им расширенную трактовку. Сам по себе анализ данных, особенно основанных на количественных оценках, часто представляет собой техническую процедуру, для выполнения которой нужно владеть соответствующими навыками статистических расчетов и придерживаться определенных правил. Интерпретация и обсуждение полученных результатов требует от исследователя определенного воображения, широты кругозора, что позволит проникнуть в суть изученного феномена. Вот почему интерпретация и обсуждение полученных результатов — это более широкая концептуальная задача, направленная на увязывание результатов с теми проблемами, которые послужили стимулом для проведения исследования.

Интерпретация и обсуждение результатов содержат принципиальные фрагменты:

— попытку понять, что означают полученные результаты с учетом результатов ранее проводившихся исследований и в свете имеющихся научных теорий;

- критическую оценку сильных и слабых сторон исследования, объективную оценку того, в какой мере результаты можно рассматривать как обоснованные аргументы предложенной интерпретации;
- рассмотрение научного и клинического значения полученных результатов, обсуждение того, как эти результаты влияют на возможное направление дальнейших исследований, насколько они применимы в клинической практике (Barker C. [et al.], 2002; Peat J. K., 2002).

6.4.1. Значение результатов

После того как исследователь убедится в достоверности полученных результатов, необходимо понять, что же они означают. Для этого их нужно сопоставить с литературными данными, которые отражают предшествующие исследования по изучаемой теме, теоретические положения или затрагивают те концептуальные модели, на которых базировалось исследование:

- В какой мере полученные в ходе исследования результаты отвечают на поставленные вопросы?

- Соответствуют ли они результатам предшествующих исследований или противоречат им, а если противоречат, то в чем возможные причины таких противоречий?

- Удовлетворяют ли результаты тем теоретическим моделям, на которых основывалось исследование, или противоречат этим моделям?

- В случае расхождений между фактическими и ожидаемыми результатами в чем заключаются возможные причины такого расхождения?

Допустимы несколько отвлеченные (однако же не слишком посторонние) рассуждения по поводу значения выявленных результатов. В англоязычной литературе это обозначается термином *speculations*, который имеет позитивную коннотацию. При этом исследователю следует предупредить читателя, что речь идет исключительно о предположениях, а не о строго установленных фактах.

Исследователю также нужно руководствоваться принципом: «Факты — наши друзья». В данном случае вряд ли будет уместно избыточное упорство и догматизм в защите собственных теорий, не подтвержденных результатами. С другой стороны, возможна попытка объяснить такое несоответствие исключительно отсутствием валидности результатов: в таком случае исследователь вместо того, чтобы скорректировать свою теорию, пытается «скорректировать» реальность. Исследование может помочь переосмыслить собственные идеи, пусть даже это вызовет неприятные эмоции. Однако если исследователь не готов к тому, чтобы пересматривать свои взгляды, в чем же тогда смысл исследования?

Наконец, исследователям следует критически оценить сильные и слабые стороны своего теоретического объяснения. Существует ли возможность объяснить полученные результаты с иных позиций, отличных от тех, которые теоретически обосновываются исследователем? Укладываются ли теоретические положения, отстаиваемые исследователем, в рамки иных принятых теоретических положений? Нередко такие рассуждения позволяют исследователю скорректировать свои теоретические положения, внести в них необходимые поправки, а это, в свою очередь, может стимулировать проведение дальнейших исследований. С практической точки зрения такое самокритичное и объективное отношение исследователя к своим данным означает, что он не должен поддаваться соблазну отбирать только те результаты исследования, которые подтверждают выдвинутую им гипотезу. В действительности результаты, отклоняющиеся от ожиданий исследователя, заслуживают такого же внимания и обсуждения, как и результаты, подтверждающие эти ожидания. Такие «негативные» результаты являются ключевыми для лучшего понимания объекта исследования, улучшения дизайна исследования, подходов к сбору данных и методов анализа.

При обсуждении результатов исследований у исследователя может и должна возникнуть потребность вновь обратиться к анализу литературных источников, для того чтобы сравнить свои результаты с тем, что было сделано другими исследователями. Только синтез чужих и собственных данных позволит автору существенно продвинуться в понимании природы исследуемого явления, поэтому в этом разделе весьма желательно дать ссылки на литературные данные. При этом совершенно не обязательно использовать лишь те источники, которые были приведены в «Обзоре литературы», в «Обсуждении» могут быть ссылки и на иные публикации.

6.4.2. Достоинства и недостатки исследования

Все исследования имеют свои сильные и слабые стороны. Задача исследователя и читателя состоит в том, чтобы определить относительный вес достоинств и недостатков исследования. Некоторые недостатки могут быть весьма тривиальными, и о них достаточно упомянуть вскользь. Другие недостатки могут быть более серьезными, но могут уравниваться (или не уравниваться) его определенными достоинствами. Серьезность недостатков исследования определяет, в какой степени можно доверять его результатам. Исследователь должен задать себе вопрос: «Есть ли какие-то проблемы, связанные с исследованием, которые могли повлиять на его результаты?» Обязанность исследователя перед самим собой и читате-

лями — открыто обозначить эти проблемы. Вспомним слова известного физика Р.Фейнмана: «То, что я собираюсь сообщить, мы никогда прямо не обсуждаем, но надеемся, что вы все вынесли это из школы: вся история научных исследований наводит на эту мысль. Поэтому стоит назвать ее сейчас со всей определенностью. Это научная честность, принцип научного мышления, соответствующий полнейшей честности, честности, доведенной до крайности. Например, если вы ставите эксперимент, вы должны сообщать обо всем, что, с вашей точки зрения, может сделать его несостоятельным. Сообщайте не только то, что подтверждает вашу правоту. Приведите все другие причины, которыми можно объяснить ваши результаты, все ваши сомнения, устраненные в ходе других экспериментов, и описания этих экспериментов, чтобы другие могли убедиться, что они действительно устранены. Если вы подозреваете, что какие-то детали могут поставить под сомнение вашу интерпретацию, — приведите их. Если что-то кажется вам неправильным или предположительно неправильным, сделайте все, что в ваших силах, чтобы в этом разобраться. Если вы создали теорию и пропагандируете ее, приводите все факты, которые с ней не согласуются, так же как и те, которые ее подтверждают» (Фейнман Р. Ф., 2001).

Исследователь должен критически оценить внутреннюю валидность исследования, о которой говорилось в предыдущих разделах. Для этого целесообразно ответить на следующие вопросы:

— Возможно ли, что на полученные результаты повлияли неучтенные факторы?

— Могут ли полученные результаты исследования быть объяснены какими-то иными переменными, которые не рассматривались и не анализировались в ходе исследования?

6.4.3. Научное и практическое значение

Научное исследование, как правило, представляет собой циклический процесс, причем собранные данные и полученные результаты лишь частично позволяют ответить на поставленные вопросы. Нередко исследование позволяет понять, что рассматриваемые вопросы могли быть сформулированы более точно; что теория, на которой базировалась гипотеза исследования, была не совсем адекватной. Часто выясняется, в чем состояли слабые стороны дизайна исследования и проведенных измерений. Кроме того, подходы, использовавшиеся в исследовании, могут оказаться весьма перспективными, что можно считать основанием для их развития и более широкого использования. Все эти выводы имеют научное значение, так как позволяют планировать последующие исследования в соответствующей области.

Не менее важным для клинико-психологических исследований является обсуждение того, какие последствия для профессиональной практической деятельности могут иметь полученные результаты. Что можно порекомендовать специалистам-практикам (врачам, клиническим психологам, педагогам и т. п.) и от чего их нужно предостеречь, опираясь на полученные результаты исследования?

Завершая подготовку текста, исследователь должен вновь критически оценить качество проведенного исследования и его изложения. Предложены важнейшие критерии такой оценки, которыми можно руководствоваться и при критической оценке исследований других авторов:

1. Литературный обзор

1.1. Обоснована ли актуальность темы?

1.2. Проводились ли ранее аналогичные исследования?

1.3. Что может добавить настоящее исследование в теорию и практику?

1.4. Относится ли проанализированная литература к теме исследования и насколько она современна?

1.5. Затрагивает ли проанализированная литература эмпирические, методологические и теоретические аспекты? Обосновывает ли она место настоящего исследования в контексте общенаучных исследований в изучаемой области?

1.6. Четко ли сформулирована цель исследования?

1.7. Четко ли сформулированы гипотеза и задачи исследования?

2. Дизайн исследования

2.1. Описаны ли дизайн исследования и процедуры, связанные с его реализацией, достаточно детально, так что у читателя есть возможность понять, что было сделано, и при желании воспроизвести исследование?

2.2. Соответствует ли дизайн исследования поставленным вопросам и сформулированным задачам?

2.3. Позволит ли дизайн исследования сделать необходимое заключение?

2.4. Достаточно ли четко описана процедура формирования выборки испытуемых?

2.5. Являются ли численность выборки и ее состав адекватными задачам исследования и процедурам анализа данных?

2.6. Удовлетворяет ли дизайн исследования соответствующим профессиональным и этическим стандартам?

3. Измерения

3.1. Как определены основные конструкты, изучаемые в ходе исследования, и как они измеряются?

3.2. Позволяют ли используемые методики измерения (в широком смысле, включая методики, основанные как на количественных, так и на качественных методах) адекватно оценить интересующие исследователя конструкты?

3.3. Адекватны ли методики тем вопросам, которые стоят перед исследователем, и сформулированным задачам исследования?

3.4. Если используются методики, основанные на количественных оценках, обеспечивают ли они приемлемые надежность и валидность измерения?

3.5. Если используются методики, основанные на качественных оценках, согласуются ли они с теоретическими взглядами исследователя?

4. Анализ

4.1. Затрагивает ли проведенный анализ каждый вопрос, стоящий перед исследователем, и каждую сформулированную задачу исследования?

4.2. Представлены ли данные четко и согласованно? (Таблицы и рисунки должны быть понятны и информативны.)

4.3. При использовании количественных оценок: корректно ли представлены обобщающие показатели (редукция данных) и корректно ли использованы статистические тесты?

4.4. При использовании качественных оценок: имеются ли примеры, какие процедуры использовались для проверки надежности анализа?

5. Интерпретация

5.1. Проводится ли интерпретация результатов в контексте тех вопросов, которые рассматривает исследование, и более широком теоретическом контексте, связанном с исследованием?

5.2. Подтверждается ли интерпретация фактическими данными? (Допустимы предположения, на что исследователь должен указать.)

5.3. Рассматриваются ли возможные альтернативные объяснения полученных результатов?

5.4. Рассматривается ли вопрос о том, насколько обобщаемы полученные результаты?

5.5. Рассматриваются ли слабые стороны исследования?

5.6. Обсуждается ли научное и практическое значение полученных результатов?

6. Стилль изложения

6.1. Доступен ли текст для восприятия? Не содержит ли он стилистических недочетов, жаргонизмов и некорректных высказываний?

6.2. Соответствует ли объем текста его содержанию и установленным требованиям? (Barker C. [et al.], 2002).

Вопросы для самоконтроля

1. Какова структура дипломной работы и научной статьи?
2. Что следует отразить в разделе «Материал и методы исследования»?
3. Какова примерная структура раздела «Результаты»?
4. Какова примерная структура раздела «Обсуждение результатов»?
5. На какие компоненты публикации следует обратить внимание для оценки ее качества?

ЗАКЛЮЧЕНИЕ

Как справедливо замечено, главная польза научного исследования «...та же, что и у розы, песни или прекрасного пейзажа, — они доставляют нам удовольствие. Каждое научное открытие выявляет новую грань в гармонии Природы для нашего пассивного наслаждения. Но наука не только “зрелищный вид спорта” — ученый активно участвует в раскрытии прекрасного. Этот тип деятельности — единственный из всех, доступных человеческому разуму, — наиболее близок к процессу творчества» (Селье Г., 1987). Хорошее исследование подобно увлекательному путешествию в мир неизведанного. Оно предполагает сборы в дорогу (подготовительный этап), прокладывание маршрута (определение дизайна исследования), само путешествие (проведение измерений и анализ данных) и осмысление увиденного (интерпретация результатов). Подобно тому как путешественник выбирает свой маршрут, исследователь формирует свой подход к проведению исследования. Это означает, что нет какого-то единого подхода к организации исследования вообще, который можно было бы признать лучшим. Выбор оптимального метода зависит от конкретных вопросов, ответы на которые хочет получить исследователь, и в этом смысле можно говорить об адекватности метода в конкретном случае. У каждого метода и методики имеются свои сильные и слабые стороны. Вот почему в данной книге отражена идея методологического плюрализма, на что должен был обратить внимание читатель. При этом следует понимать, что методологический плюрализм не означает методологической анархии. Напротив, существуют определенные методологические нормы и правила, которых целесообразно придерживаться при планировании и организации исследования, и в книге они последовательно изложены. Соблюдение этих норм и правил призвано защитить как самого исследователя, так и тех, кому будут адресованы результаты исследования, от возможных заблуждений, связанных с выводами, не подкрепленными фактическими данными. Знание принципов, на которых основан верный выбор метода, является существенным, хотя и не единственным условием успешного выполнения исследования. Однако хочется надеяться, что знакомство с этими принципами поможет сделать исследователю обдуманый выбор, который позволит максимально оптимизировать исследование или, по крайней мере, пойти на осознанные и разумные компромиссы при планировании и проведении исследования. Как явствует из изложения материала, подобные компромиссы практически всегда неизбежны.

Если можно с уверенностью утверждать, что нет одного-единственного лучшего способа провести исследование, то с не меньшей

уверенностью можно говорить и о том, что всегда имеются плохие способы. Вот почему принципиально важно уметь критически взглянуть как на свое собственное исследование, так и на исследования коллег, чтобы понять, насколько они хороши. В книге представлены важнейшие концептуальные сведения, которые должны помочь читателю выступать как в роли «добросовестного производителя», так и «культурного потребителя» исследовательской продукции. Чем лучше клинический психолог знаком с методологией исследования, тем легче ему распознать проблемы, связанные с исследованием. Однако критичность по отношению к исследованию не означает желания непременно найти в нем изъяны. Любое исследование всегда легче критиковать — гораздо сложнее увидеть содержащиеся в нем конструктивные элементы. При вынесении суждения об исследовании в целом нужно иметь в виду, что ряд сильных сторон исследования может компенсировать имеющиеся в нем методологические недостатки. Если в исследовании и имеются методологические дефекты, то следует не просто констатировать их наличие, а понять, в какой степени они могли повлиять на результаты и сделанные на их основании выводы. Так, некоторые технические дефекты при выполнении ряда процедур исследования могут практически не отразиться на результатах. При этом надо помнить, что чем более новаторскими являются тема исследования, используемые методы или методики, тем более снисходительным следует быть в отношении неизбежных в таких ситуациях огрехов.

Знание принципов методологии имеет принципиальное значение не только для исследователей-теоретиков. Исследования помогают обогатить практику сведениями, касающимися факторов риска ряда патологических состояний, подходами к профилактике, диагностике и лечению ряда расстройств. Кроме того, практические наблюдения часто являются источником новых идей и предостерегают исследователей от излишне упрощенного подхода к изучению поведения человека. Хотя практикующие клинические психологи и не столь сильно вовлечены в исследовательскую деятельность, они тем не менее должны быть достаточно осведомлены об основных методах исследований, чтобы быть готовыми понять и осмыслить суть той или иной работы. Дизайн и масштаб ряда исследований таков, что вряд ли практикующие специалисты могут провести их только собственными силами (например, рандомизированные контролируемые испытания). Вряд ли можно предполагать, что всякий практик обладает всеми ресурсами, которые необходимы для организации и проведения исследования. Однако прототипы некоторых дизайнов исследований специалист-практик реализует в своей повседневной деятельности, например динамическое наблюдение за одним или несколькими па-

циентами во время лечения, и знания методологических принципов организации подобных исследований может помочь ему правильно оценить результаты своей работы. Кроме того, современная тенденция диктует необходимость все больше вовлекать практиков в решение исследовательских задач, организовывать их сотрудничество с научными работниками, формировать исследовательские группы. Для того чтобы понимать друг друга, практикам и теоретикам следует «разговаривать на одном языке».

В настоящее время исследования не следует рассматривать как удел избранных. Приведенные сведения о методологии исследования призваны убедить читателя в том, что существует большое многообразие видов и дизайнов исследований, и каждый потенциальный исследователь может выбрать для себя жанр и стиль исследования по своему вкусу, возможностям и стилю мышления. Тем не менее не все так просто. Исследование требует времени и усилий, серьезного интеллектуального напряжения, да и соблюдение строгости мышления — нелегкая задача. Однако хочется надеяться, что радость познания, связанная с проведением исследования, должна вдохновить читателя. Продолжая аналогию с путешествием, можно сказать, что знание методологии исследования позволит клиническому психологу не заблудиться самому и не сбить с пути тех, кто следует за ним.

Приложение 1. Таблица случайных чисел (Вевер В. Н., 1968)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52686	19174	39615	99505
	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
	42187	93093	08243	61680	07856	16376	39440	53537	71341	57004	00849	74917	93758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05385	56941
10	85475	36857	53342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07058	97628	33787	09998	42698	08691	76988	13602	51851	46104	88916	19509	25625	58104
	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
	54164	58492	22421	74103	47070	25306	76468	26384	53151	06646	21524	15227	96909	44592

32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720
	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666
	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428
	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096
	05366	04213	25869	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07844
	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028
	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919
	69011	65795	95876	55293	18988	27354	26575	40825	40801	59920	29841	80150	12777
	25976	57943	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917
	91567	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656
	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83517	36103
	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562
	92157	89834	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490
	98427	07523	33362	64270	01638	92477	69699	98420	04880	45585	46565	40102	46880
	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775
	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102
	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541
45	76072	29516	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697
	90725	52210	83974	23992	65831	38857	50490	83765	55657	14361	31720	57375	56228
	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726
	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547
	95012	68379	93526	70765	10592	04542	76463	54328	02349	17247	28865	14777	62730
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261
													85653

	15	16	17	18	19	20	21	22	23	24	25	26	27	28
51	16408	81399	04153	53381	79401	21438	83035	92350	36693	31238	59649	91754	72772	02338
	18629	31953	05520	91962	04739	13092	97662	24822	94730	06496	35090	04822	86774	98289
	73115	35101	47498	86737	99016	71060	88824	71013	18735	20286	23153	72924	35165	43040
	57491	16703	23167	49323	45921	33132	12544	41035	80780	45393	44812	12515	98931	91202
55	30405	83946	23792	14422	15059	45799	22716	19792	09983	74353	68668	30429	70735	25499
	16631	35006	85900	98275	32388	52390	16815	69298	82732	38480	73817	32523	41961	44437
	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052	91491	60383	19746
	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318	29686	03387	59846
	31624	76384	17403	53363	44167	64486	64758	75366	76554	31601	12614	33072	60332	92325
60	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806	08930	35001	87820
	03931	33309	57047	74211	63445	17361	62825	39908	05607	91284	68833	25570	38818	46920
	74426	33278	43972	10119	89917	15665	52872	73823	73144	88662	88970	74492	51805	93378
	09066	00903	20795	95452	92648	45454	09552	88815	16553	51125	79375	97596	16296	66092
	42238	12426	87025	14267	20979	04508	64535	31355	86064	29472	47689	05974	52468	16834
65	16153	08002	26504	41744	81959	65642	74240	56302	00033	67107	77510	70625	28725	34191
	21457	40742	29820	96783	29400	21840	15035	34537	33310	06116	95240	15957	16572	06004
	21581	57802	02050	89728	17937	37621	47075	42080	97403	48626	68995	43805	33386	21597
	55612	78095	83197	33732	05810	24813	86902	60397	16489	03264	88525	42786	05269	92532
	44657	66999	99324	51281	84463	60563	79312	93454	68876	25471	93911	25650	12682	73572
70	91340	84979	46949	81973	37949	61023	43997	15263	80644	43942	89203	71795	99533	50501
	91227	21199	31935	27022	84067	05462	35216	14486	29891	68607	41867	14951	91696	85065
	50001	38140	66321	19924	72163	09538	12151	06878	91903	18749	34405	56087	82790	70925
	65390	05224	72958	28609	81406	39147	25549	48542	42627	45233	57202	94617	23772	07896
	27504	98131	83944	41575	10573	03619	64482	73923	36152	05184	94142	25299	84387	34925

75	37169	94851	39117	89632	00959	16487	65536	49071	39782	17095	02330	74301	00275	48280
	11508	70225	51111	38351	19444	66499	71945	05422	13442	78675	84081	66938	93654	59894
	37449	30362	06694	54690	04052	53115	62757	95348	78662	11163	81651	50245	34971	52924
	46515	70331	85922	38329	57015	15765	97161	17869	45349	61796	66345	81073	49106	79860
	30986	81223	42416	58353	21532	30502	32305	86482	05174	07901	54339	58861	74818	46942
80	63798	64995	46583	09785	44160	78128	83991	42865	92520	83531	80377	35909	81250	54238
	82486	84846	99254	67632	43218	50076	21361	64816	51202	88124	41870	52689	51275	83556
	21885	32906	92431	09060	64297	51674	64126	62570	26123	05155	59194	52799	28225	85762
	60336	98782	07408	53458	13564	59089	26445	29789	85205	41001	12535	12133	14645	23541
	43937	46891	24010	25560	86355	33941	25786	54990	71899	15475	95434	98227	21824	19585
85	97656	63175	89303	16275	07100	92063	21942	18611	47348	20203	18534	03862	78095	50136
	03299	01221	05418	38982	55758	92237	26759	86367	21216	98442	08303	56613	91511	75928
	79626	06486	03574	17668	07785	76020	79924	25651	88325	88428	85076	72811	22717	50585
	85636	68335	47539	03129	65651	11977	02510	26113	99447	68645	34327	15152	55230	93448
	18039	14367	61337	06177	12143	46609	32989	74014	64708	00533	35398	58408	13261	47908
90	08362	15656	60627	36478	65648	16764	53412	09013	07832	41574	17639	82163	60859	75567
	79556	29068	04142	16268	15387	12856	66227	38358	22478	73373	88732	09443	82558	05250
	92608	82674	27072	32534	17075	27698	98204	63863	11951	34648	88022	56148	34925	57031
	23982	25835	40055	67006	12293	02753	14827	23235	35071	99704	37543	11601	35503	85171
	09915	96306	05908	97901	28395	14186	00821	80703	70426	75647	76310	88717	37890	40129
95	59037	33300	26695	62247	69927	76123	50842	43834	86654	70959	79725	93872	28117	19233
	42488	78077	69882	61657	34136	79180	97526	43092	04098	73571	80799	76536	71255	64239
	46764	86273	63003	93017	31204	36692	40202	35275	57306	55543	53203	18098	47625	88684
	03237	45430	55417	63282	90816	17349	88298	90183	36600	78406	06216	95787	42579	90730
	86591	81482	52667	61582	14972	90053	89534	76036	49199	43716	97548	04379	46370	28672
100	38534	01715	94964	87288	65680	43772	39560	12916	86537	62738	19636	51132	25739	56947

**Приложение 2. Критические значения *t*-распределения
Стьюдента для двустороннего критерия (Woodward M., 2014)**

<i>P</i>	50	20	10	5	2	1	0,2	0,1
<i>v</i> = 1	1,00	3,08	6,31	12,7	31,8	83,7	318	637
2	0,82	1,89	2,92	4,30	6,96	9,92	22,3	81,6
3	0,76	1,69	2,35	3,18	4,54	5,84	10,2	12,9
4	0,74	1,53	2,13	2,78	3,75	4,60	7,17	8,61
5	0,78	1,48	2,02	2,57	3,36	4,03	5,89	6,87
6	0,72	1,44	1,94	2,45	3,14	3,71	5,21	5,96
7	0,71	1,42	1,89	2,36	8,00	3,50	4,79	5,41
8	0,71	1,40	1,86	2,31	2,90	3,36	4,50	5,04
9	0,70	1,38	1,83	2,26	2,62	3,25	4,30	4,78
10	0,70	1,37	1,81	2,23	2,76	3,17	4,14	4,69
11	0,70	1,36	1,80	2,20	2,72	3,11	4,03	4,44
12	0,70	1,36	1,78	2,18	2,68	3,05	3,93	4,32
13	0,69	1,36	1,77	2,16	2,65	3,01	3,85	4,22
14	0,69	1,36	1,76	2,14	2,62	2,98	3,79	4,14
15	0,69	1,34	1,75	2,13	2,60	2,95	8,73	4,07
16	0,69	1,34	1,75	2,12	2,58	2,92	3,69	4,01
17	0,69	1,33	1,74	2,11	2,57	2,90	3,65	3,96
18	0,69	1,33	1,73	2,10	2,55	2,88	3,61	3,92
19	0,69	1,33	1,73	2,09	2,54	2,86	3,58	3,88
20	0,69	1,32	1,72	2,09	2,53	2,85	3,55	3,85
22	0,69	1,32	1,72	2,07	2,51	2,82	3,51	3,79
24	0,68	1,32	1,71	2,06	2,49	2,80	3,47	3,75
26	0,68	1,32	1,71	2,06	2,48	2,78	3,44	3,71
28	0,68	1,31	1,70	2,05	2,47	2,76	3,41	3,67
30	0,63	1,31	1,70	2,04	2,46	2,75	3,39	3,65
35	0,68	1,31	1,69	2,03	2,44	2,72	3,34	3,59
40	0,88	1,30	1,68	2,02	2,42	2,70	3,31	3,55
45	0,88	1,30	1,68	2,01	2,41	2,60	3,28	3,52
50	0,68	1,30	1,68	2,01	2,40	2,68	3,26	3,50
55	0,68	1,30	1,67	2,00	2,40	2,67	3,25	3,48
60	0,68	1,30	1,67	2,00	2,39	2,66	3,23	3,46
∞	0,67	1,28	1,64	1,96	2,33	2,56	3,09	3,29

**Приложение 3. Критические значения распределения χ^2
(Woodward M., 2014)**

<i>P</i>	97,5	95	50	10	5	2,5	1	0.1
v = 1	0,000982	0,00393	0,45	2,71	3,84	5,02	6,64	10,8
2	0,0506	0,103	1,39	4,61	5,99	7,38	9,21	13,8
3	0,216	0,352	2,37	6,25	7,81	9,35	11,3	16,3
4	0,484	0,711	3,36	7,78	9,49	11,1	13,3	18,5
5	0,831	1,15	4,35	9,24	11,1	12,8	15,1	20,5
6	1,24	1,64	5,35	10,6	12,6	14,5	16,8	22,5
7	1,69	2,17	6,35	12,0	14,1	16,0	18,5	24,3
8	2,18	2,73	7,34	13,4	15,5	17,5	20,1	26,1
9	2,70	3,33	8,34	14,7	16,9	19,0	21,7	27,9
10	3,25	3,94	9,34	16,0	18,3	20,5	23,2	29,6
11	3,82	4,57	10,3	17,3	19,7	21,9	24,7	31,3
12	4,40	5,23	11,3	18,5	21,0	23,3	26,2	32,9
13	5,01	5,89	12,3	19,8	22,4	24,7	27,7	34,5
14	5,63	6,57	13,3	21,1	23,7	26,1	29,1	36,1
15	6,26	7,26	14,3	23,3	25,0	27,5	30,6	37,7
16	6,91	7,96	15,3	23,5	26,3	28,8	32,0	39,3
17	7,56	8,67	16,3	24,8	27,6	30,2	33,4	40,8
18	8,23	9,39	17,3	26,0	28,9	31,5	34,8	42,3
19	8,91	10,1	18,3	27,2	30,1	32,9	36,2	43,8
20	9,59	10,9	19,3	28,4	31,4	34,2	37,6	45,3
22	11,0	12,3	21,3	30,8	33,9	36,8	40,3	48,3
24	12,4	13,9	23,3	33,2	36,4	39,4	43,0	51,2
26	13,8	16,4	25,3	35,6	38,9	41,9	45,6	54,1
28	15,3	16,9	27,3	37,9	41,3	44,5	48,3	56,9
30	16,8	18,5	29,3	40,3	43,8	47,0	50,9	59,7
35	20,6	22,5	34,3	46,1	49,8	53,2	57,3	66,6
40	24,4	26,5	39,3	51,8	55,8	59,3	63,7	73,4
45	28,4	30,6	44,3	57,5	61,7	65,4	70,0	80,1
50	32,4	34,8	49,3	63,2	67,5	71,4	76,2	86,7
55	36,4	39,0	54,3	68,8	73,3	77,4	82,3	93,2
60	40,5	43,2	59,3	74,4	79,1	83,3	88,4	99,7

ЛИТЕРАТУРА

Баскаков А. Я., Туленков Н. В. Методология научного исследования : учеб. пособие. — Киев : МАУП, 2004. — 216 с.

Власов В. В. Введение в доказательную медицину. — М. : Медиа Сфера, 2001. — 392 с.

Власов В. В. Эпидемиология. — М. : ГЭОТАР-Медиа, 2004. — 462 с.

Волков Б. С., Волкова Н. В., Губанов А. В. Методология и методы психологического исследования. — М. : Академический проект ; Фонд «Мир», 2005. — 352 с.

Голев С. В., Голева О. С. Методология и методы научного исследования. Методика психологического исследования. Ч. 1. — Херсон : ОМУРЧ «Украина» ХФ, 2010а. — 213 с.

Голев С. В., Голева О. С. Методология и методы научного исследования. Ч. 2 : Методика психологического исследования. — Херсон : ОМУРЧ «Украина» ХФ, 2010б. — 173 с.

Гудвин Д. Исследование в психологии: методы и планирование : пер. с англ. — СПб. : Питер, 2004. — 558 с.

Карвасарский Б. Д. Клиническая психология : учеб. для вузов. — СПб. : Питер, 2011. — 864 с.

Кельмансон И. А. Клиническая эпидемиология в педиатрии. — СПб. : Сотис, 2002. — 208 с.

Кельмансон И. А. Принципы доказательной педиатрии. — СПб. : Феникс, 2004. — 240 с.

Кельмансон И. А. Мета-анализ в педиатрии и детской клинической психологии. Теория и практика систематических обзоров. — Saarbrücken : LAP Lambert Academic Publishing, 2011. — 208 с.

Кельмансон И. А. Перинатология и перинатальная психология. — СПб. : СпецЛит, 2015. — 344 с.

Максименков А. Н. Лекции по методике научно-исследовательской работы. — Л. : ВМедА, 1972. — 108 с.

Марьянович А. Т., Князькин И. В. Диссертация: Инструкция по подготовке и защите. — М. ; СПб. : Астрель, 2009. — 231 с.

Мелихов О. Г. Клинические исследования. — М. : Атмосфера, 2003. — 200 с.

Менделевич В. Д. Клиническая и медицинская психология : практическое руководство. — М. : МЕДпресс, 2001. — 592 с.

Наследов А. Д. Математические методы психологического исследования. Анализ и интерпретация данных. — СПб. : Речь, 2004. — 392 с.

Новик А. А., Ионова Т. И. Руководство по исследованию качества жизни в медицине. — М. : ЗАО «ОЛМА Медиа Групп», 2007. — 320 с.

Перре М., Бауманн У. Клиническая психология : пер. с нем. А. Желнина, Е. Михалевич, Н. Римицан [и др.]. — СПб. : Питер, 2006. — 1312 с.

Покровский В. П., Брико Н. И. Общая эпидемиология с основами доказательной медицины: руководство к практическим занятиям : учеб. пос. — М. : ГЭОТАР-Медиа, 2012. — 496 с.

Поппер К. Логика научного исследования / пер. с англ. — М. : Республика, 2004. — 447 с.

Селье Г. От мечты к открытию: Как стать ученым : пер. с англ. Н. И. Войскунской. — М. : Прогресс, 1987. — 368 с.

Сидоренко Е. В. Методы математической обработки в психологии. — СПб. : Речь, 2007. — 350 с.

Соложенкин В. В. Избранные лекции по психиатрии для врачей-кардиологов : учеб. пос. — Бишкек : Изд-во КРСУ, 2011. — 134 с.

Уайтхед А. Избранные работы по философии : пер. с англ. — М. : Прогресс, 1990. — 718 с.

Фейнман Р. Ф. «Вы, конечно, шутите, мистер Фейнман!» : пер. с англ. Н. А. Зубченко, О. Л. Тиходеевой, М. Шифмана. — М. : НИЦ «Регулярная и хаотическая динамика», 2001. — 336 с.

Abramson J. H. WINPEPI updated: computer programs for epidemiologists, and their teaching potential // *Epidemiol Perspect Innov.* — 2011. — Vol. 8, № 1. — P. 1–9.

Altman D. G. Practical statistics for medical research. — Boca Raton, Fla. : Chapman & Hall/CRC, 1999. — 611 p.

Alzheimer's disease facts and figures // Alzheimer's & Dementia: The Journal of the Alzheimer's Association. — 2012. — Vol. 8, № 2. — P. 131–168.

Armitage P., Berry G., Matthews J. N. S. Statistical methods in medical research. — Malden, MA: Blackwell Science, 2001. — 817 p.

Aveline M., Shapiro D. Research foundations for psychotherapy practice. — Chichester ; New York : J. Wiley in association with the Mental Health Foundation, 1995. — 332 p.

Bailar J. C., Hoaglin D. C. Medical uses of statistics. — Hoboken, N. J. : John Wiley & Sons, 2009. — 491 p.

Baker T. B., McFall R. M., Shoham V. Current Status and Future Prospects of Clinical Psychology: Toward a Scientifically Principled Approach to Mental and Behavioral Health Care // *Psychol Sci Public Interest.* — 2008. — Vol. 9, № 2. — P. 67–103.

Barker C., Pistrang N., Elliott R. Research methods in clinical psychology : an introduction for students and practitioners. — New York : J. Wiley, 2002. — 284 p.

Baron R. M., Kenny D. A. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations // *J. Pers. Soc. Psychol.* — 1986. — Vol. 51, № 6. — P. 1173–1182.

Bax L., Yu L.-M., Ikeda N. [et al.]. Development and validation of MIX: comprehensive free software for meta-analysis of causal research

data // BMC Medical Research Methodology. — 2006. — Vol. 6, № 50. — P. 1—11.

Beyer W. H. Handbook of tables for probability and statistics. — Cleveland : Chemical Rubber Co, 1968. — 656 p.

Bolhuis S., Voeten M. Towards self-directed learning in secondary schools: What do teachers do? // Teaching and Teacher Education. — 2001. — Vol. 17, № 7. — P. 837—855.

Borenstein M., Hedges L., Higgins J. [et al.]. Comprehensive Meta-analysis Version 2. — Englewood, N. J. : Biostat, 2005. — 104 p.

Borenstein M., Hedges L. V., Higgins J. P. T. [et al.]. Introduction to meta-analysis. — Oxford : Wiley, 2009. — 421 p.

Brazelton T. B., Nugent J. K. Neonatal behavioral assessment scale. — London : Mac Keith Press, 2011. — 200 p.

Campbell D. T., Stanley J. C. Experimental and quasi-experimental designs for research. — Chicago : R. McNally, 1963. — 84 p.

Cardinet J., Johnson S., Pini G. Applying generalizability theory using EduG. — New York : Routledge, 2010. — 215 p.

Choi S. K., Kim J. J., Park Y. G. [et al.]. The simplified Edinburgh Postnatal Depression Scale (EPDS) for antenatal depression: is it a valid measure for pre-screening? // Int J. Med. Sci. — 2012. — Vol. 9, № 1. — P. 40—46.

Chow S.-C., Shao J., Wang H. Sample size calculations in clinical research. — Boca Raton : Chapman & Hall/CRC, 2008. — 465 p.

Cohen B. H., Lea R. B. Essentials of statistics for the social and behavioral sciences. — Hoboken, N. J. : Wiley, 2004. — 289 p.

Cook T. D., Campbell D. T. Quasi-experimentation : design & analysis issues for field settings. — Chicago : Rand McNally College Pub. Co., 1979. — 405 p.

Cooper B., Eastwood M. R., World Psychiatric Association. Section of Epidemiology and Community Psychiatry. Primary health care and psychiatric epidemiology. — London ; New York : Tavistock/Routledge, 1992. — 357 p.

Cronbach L. J. The Dependability of behavioral measurements: theory of generalizability for scores and profiles. — New York : Wiley, 1972. — 410 p.

Cunningham H., Maynard D., Bontcheva K. [et al.]. Developing Language Processing Components with GATE Version 8. — Sheffield : The University of Sheffield, Department of Computer Science, 2015. — 681 p.

D'Agostino R. B., Weintraub M. Meta-analysis: a method for synthesizing research // Clin. Pharmacol. Ther. — 1995. — Vol. 58. — P. 605—616.

Damer T. E. Attacking faulty reasoning. A Practical Guide to Fallacy-Free Arguments. — Belmont, CA: Wadsworth Cengage Learning, 2009. — 234 p.

Dean A. G., Arner T. G., Sunki G. G. [et al.]. Epi Info™, a database and statistics program for public health professionals. — Atlanta, GA, USA : CDC, 2011 p.

DeMars C. Item response theory. — Oxford ; New York : Oxford University Press, 2010. — 131 p.

Fagard R. H., Staessen J. A., Thijs L. Advantages and disadvantages of the meta-analysis approach // J. Hypertension. — 1996. — Vol. 14 (S2). — P. S9—S13.

Faul F., Erdfelder E., Lang A. G. [et al.]. Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences // Behav Res Methods. — 2007. — Vol. 39, № 2. — P. 175—191.

Ferrari A. J., Charlson F. J., Norman R. E. [et al.]. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010 // PLoS Med. — 2013. — Vol. 10, № 11. — P. e1001547.

Flleiss J. L., Levin B., Paik M. C. Statistical methods for rates and proportions. — Hoboken, N. J. : J. Wiley, 2003. — 760 p.

Franklin S., Walker C. Survey methods and practices. — Ottawa : Statistics Canada, 2010. — 396 p.

Glass G. V. Primary, secondary and meta-analysis of research // Educat. Res. — 1976. — Vol. 5. — P. 3—8.

Gray G. E. Concise guide to evidence-based psychiatry. — Washington, D. C. : American Psychiatric Pub., 2004. — 237 p.

Hayes A. F. Introduction to mediation, moderation, and conditional process analysis : a regression-based approach. — New York ; London : The Guilford Press, 2013. — 507 p.

Haynes S. N., O'Brien W. H. Principles and practice of behavioral assessment. — New York ; London : Kluwer Academic, 2000. — 348 p.

Helsel D. R., Hirsch R. M. Statistical methods in water resources. — Amsterdam ; New York : Elsevier, 1992. — 522 p.

Hennekens C. H., Buring J. E., Mayrent S. L. Epidemiology in medicine. — Boston : Little, Brown, 1987. — 383 p.

Higgins J. P. T., Green S., Cochrane Collaboration. Cochrane handbook for systematic reviews of interventions. — Chichester; Hoboken, N. J. : Wiley-Blackwell, 2008. — 649 p.

Hill A. B. The Environment and Disease: Association or Causation? // Proc R Soc Med. — 1965. — Vol. 58. — P. 295—300.

Hollander M., Wolfe D. A., Chicken E. Nonparametric statistical methods, 2014. — 819 p.

Howell D. C. Statistical methods for psychology. — Belmont, CA : Wadsworth Cengage Learning, 2013. — 770 p.

Huber G. L., Gürtler L. AQUAD 7 Manual: The analysis of qualitative data. — Tübingen : Softwarevertrieb Günter Huber, 2015. — 203 p.

Hunsley J., Lee C. M. Introduction to clinical psychology : an evidence-based approach. — Hoboken, N. J. : John Wiley & Sons, 2010. — 556 p.

Illum N. O., Gradel K. O. The Nordic Five to Fifteen questionnaire could provide the basis for a common neurological disability variable // *Acta Paediatrica*. — 2014. — Vol. 103, № 8. — P. 879—885.

Kelmanson I. A. Maternal smoking during pregnancy and sleep problems in 2-month-old infants // *Somnologie — Schlafforschung und Schlafmedizin*. — 2009. — Vol. 13, № 4. — P. 244—250.

Kelmanson I. A. Disturbed sleep and emotional reactivity in 5-year-old children // *Somnologie — Schlafforschung und Schlafmedizin*. — 2013a. — Vol. 17, № 1. — P. 49—56.

Kelmanson I. A. Emotional and behavioural features of preschool children born by Caesarean deliveries at maternal request // *Eur. J. Dev. Psychol.* — 2013b. — Vol. 10, № 6. — P. 676—690.

Kelmanson I. A. Recurrent respiratory infections and psychological problems in junior school children // *Early Child Development and Care*. — 2015. — Vol. 185, № 9. — P. 1437—1451.

Kim H. J., Fay M. P., Feuer E. J. [et al.]. Permutation tests for joinpoint regression with applications to cancer rates // *Stat. Med.* — 2000. — Vol. 19, № 3. — P. 335—351.

Kline R. B. Principles and practice of structural equation modeling. — New York : Guilford Press, 2011. — 427 p.

Kline R. B. Beyond significance testing: statistics reform in the behavioral sciences. — Washington, DC: American Psychological Association, 2013. — 349 p.

Kulinskaya E., Morgenthaler S., Staudte R. G. Meta analysis: a guide to calibrating and combining statistical evidence. — Chichester ; Hoboken; N. J. : John Wiley & Sons, 2008. — 260 p.

L'Abate L. Toward a science of clinical psychology: laboratory evaluations and interventions. — New York : Nova Science, 2008. — 414 p.

Leeuw E. D., Hox J. J., Dillman D. A. International handbook of survey methodology. — New York ; London : Lawrence Erlbaum Associates, 2008. — 549 p.

Littell J. H., Corcoran J., Pillai V. K. Systematic reviews and meta-analysis. — Oxford ; New York : Oxford University Press, 2008. — 202 p.

Lowe H. J., Barnett G. O. Understanding and using the medical subject headings (mesh) vocabulary to perform literature searches // *JAMA*. — 1994. — Vol. 271. — P. 1103—1108.

McFall R. M. Manifesto for a science of clinical psychology // *The Clinical Psychologist*. — 1991. — Vol. 44, № 6. — P. 75—88.

Meltzoff J. Critical thinking about research : psychology and related fields. — Washington, DC: American Psychological Association, 1998. — 297 p.

Meltzoff J., Kornreich M. Research in psychotherapy. — New Brunswick : AldineTransaction, 2008. — 561 p.

Mezzich J. E., Jorge M. R., Salloum I. M. Psychiatric epidemiology: assessment concepts and methods. — Baltimore : The Johns Hopkins University Press, 1994. — 612 p.

Morgenstern H., Thomas D. Principles of study design in environmental epidemiology // *Environ Health Perspect.* — 1993. — Vol. 101, Suppl. 4. — P. 23–38.

Morley S., Adams M. Some simple statistical tests for exploring single-case time-series data // *Br. J. Clin. Psychol.* — 1989. — Vol. 28, Pt 1. — P. 1–18.

Newby P. Research methods for education. — Oxon ; New York : Routledge, 2014. — 683 p.

Nikles J., Mitchell G. The essential guide to N-of-1 trials in health. — Dordrecht ; Heidelberg ; New York ; London : Springer, 2015. — 245 p.

Nunnally J. C., Bernstein I. H. Psychometric theory. — New York : McGraw-Hill, 1994. — 752 p.

Patton M. Q. Essentials of utilization-focused evaluation. — Los Angeles; Calif. : SAGE, 2012. — 461 p.

Peat J. K. Scientific writing: easy when you know how. — London : BMJ Books, 2002. — 292 p.

Peat J. K., Barton B. Medical statistics: a guide to SPSS, data analysis, and critical appraisal. — Chichester, UK : John Wiley & Sons Ltd., 2014. — 392 p.

Pompili M., Tatarelli R. Evidence-based practice in suicidology: a source book. — Cambridge, Mass. : Hogrefe, 2011. — 380 p.

Rau R. Seasonality in human mortality : a demographic approach. — Berlin; New York : Springer, 2007. — 214 p.

Reddy M. V. Statistical methods in psychiatry research and SPSS. — Toronto ; New Jersey : CRC Press, 2015. — 336 p.

Retamal C. P., Humphreys D. Occurrence of suicide and seasonal variation // *Rev. Saúde Pública.* — 1998. — Vol. 32, № 5. — P. 408–412.

Rossi P. H., Lipsey M. W., Freeman H. E. Evaluation: a systematic approach. — Thousand Oaks, CA: Sage, 2004. — 470 p.

Rothman K. J., Greenland S., Lash T. L. Modern epidemiology. — Philadelphia : Wolters Kluwer Health / Lippincott Williams & Wilkins, 2008. — 758 p.

Sargent J. D., Heatherton T. F. Comparison of trends for adolescent smoking and smoking in movies, 1990–2007 // *JAMA.* — 2009. — Vol. 301, № 21. — P. 2211–2213.

Schmidt T. EXMARaLDA Partitur-Editor. Handbuch. Version 1.5.1. – Hamburg : University of Hamburg, 2011. – 123 p.

Schüch L. CATMA user manual for CATMA 3.0. – Hamburg : University of Hamburg, 2010. – 55 p.

Senn S. Cross-over trials in clinical research. – Chichester ; New York : J. Wiley, 2002. – 345 p.

Shadish W. R., Cook T. D., Campbell D. T. Experimental and quasi-experimental designs for generalized causal inference. – Boston : Houghton Mifflin, 2001. – 623 p.

Shapiro D. A. Outcome research // Behavioural and mental health research: a handbook of skills and methods / Eds. G. Parry, F. N. Watts. – Hove : Erlbaum (UK) Taylor & Francis, 1996. – P. 201–228.

Shavelson R. J., Webb N. M. Generalizability theory: a primer. – Newbury Park, Calif. : Sage Publications, 1991. – 137 p.

Sloane R. B. Psychotherapy versus behavior therapy. – Cambridge, Mass. ; London : Harvard University Press, 1975. – 256 p.

Standen A. Science is a sacred cow. – New York : Dutton, 1950. – 221 p.

Statistical Research and Applications Branch, National Cancer Institute. Joinpoint Regression Program, version 4.2.0.2. – Bethesda, MD : Statistical Research and Applications Branch, National Cancer Institute, 2015. – 112 p.

Straus S. E. Evidence-based medicine: how to practice and teach it. – Edinburgh : Elsevier Churchill Livingstone, 2011. – 299 p.

Susser E. S. Psychiatric epidemiology: searching for the causes of mental disorders. – Oxford ; New York : Oxford University Press, 2006. – 516 p.

Szklo M., Nieto F. J. Epidemiology: beyond the basics. – Sudbury, Mass. : Jones and Bartlett Publishers, 2007. – 489 p.

Taylor C. B. How to practice evidence-based psychiatry: basic principles and case studies. – Washington, DC : American Psychiatric Pub., 2010. – 368 p.

Taylor S. J., Bogdan R. Introduction to qualitative research methods: a guidebook and resource. – New York : Wiley, 1998. – 337 p.

Townsend J. Economics of tobacco use and control // The tobacco epidemic / Ed. R. Loddenkemper, M. Kreuter. – Basel : Karger, 2015. – P. 121–135.

Tsuang M. T., Tohen M., Jones P. B. Textbook of psychiatric epidemiology. – Chichester, West Sussex ; Hoboken, N. J. : Wiley-Blackwell, 2011. – 646 p.

Woodward M. Epidemiology: study design and data analysis. – Boca Raton: Chapman & Hall/CRC Press, Taylor & Francis Group, 2014. – 832 p.

Zamora J., Abaira V., Muriel A. [et al.]. Meta-DiSc: a software for meta-analysis of test accuracy data // BMC Medical Research Methodology. – 2006. – Vol. 6, № 31. – P. 1–12.

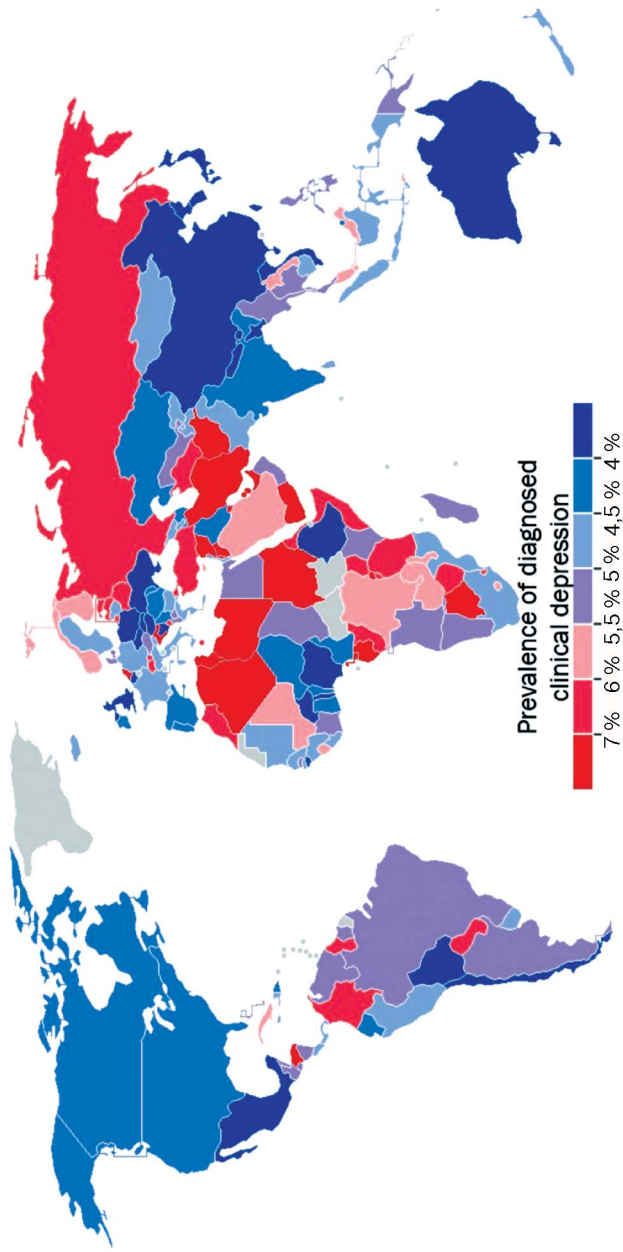


Рис. 4. Картографическое представление сведений о распространенности депрессивных расстройств в различных регионах мира по состоянию на 2010 г. (Ferrari A. J. [et al.], 2013):

красным оттенкам соответствуют более высокие показатели, синим — более низкие

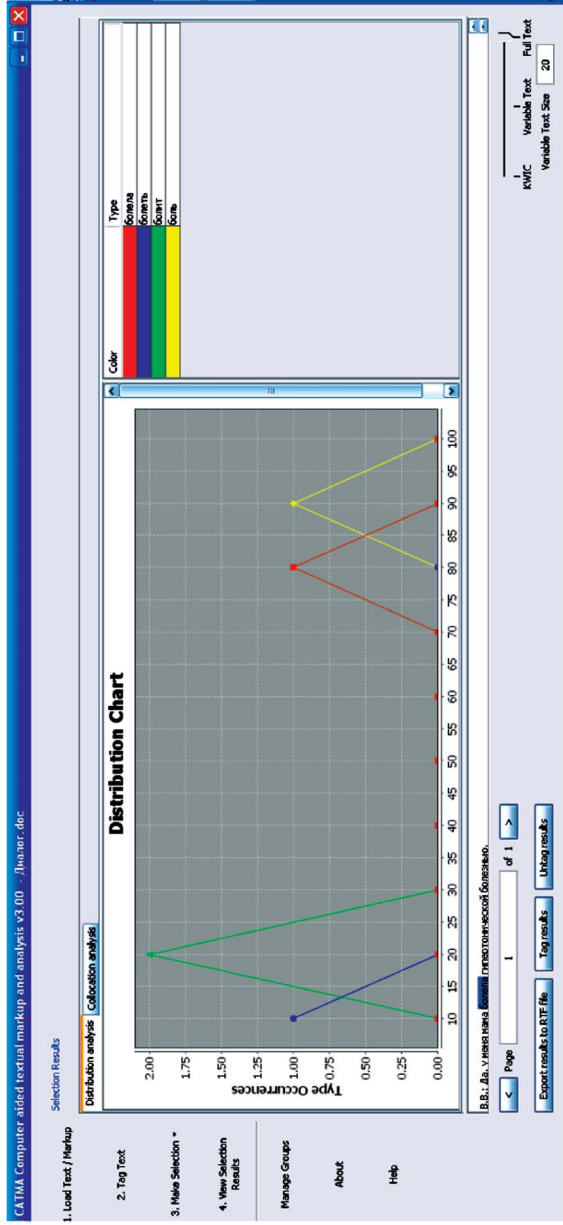


Рис. 38. Дискурсивный анализ диалога врача и пациента

СЭМА Computer aided textual markup and analysis v3.00 - /диалог.doc

Selection Results

Distribution analysis Collocation analysis

2. Tag Text

3. Make Selection *

4. View Selection Results

	Collocation frequency	Type frequency	z-score
criticism		2	4.011
если		2	4.011
меня		4	3.944
у		4	3.004
привон		1	2.834
показывает		1	2.834
незд		1	2.834
объекта		1	2.834
поскольку		1	2.834
этой		1	2.834
начинает		1	2.834
заключать		1	2.834

Доктор (далее Д.): Ваши жалобы?

Валерий Владимирович, игравший пациента (далее – В.В.):

В последнее время я замечаю следующие вещи: негод, даже не пишу отчеты, хотя это можно связать с погодой, у меня начинает болеть голова, я никак не могу спать, давление невысокое (140 на 90), но голова болит, я стал замечать, что если у меня болит в этой области (показывает), то принимаю аспирин, но пока заметного улучшения не вижу. А голова мне становится легче, я стал чувствовать себя лучше, но одновременно у меня появляется какой-то дискомфорт в голове, я могу извергать дымление, и оно оказывается 170-180.

В. В. Владимирович, В.В.

Page 1 of 1

Export results to RTF file Tag results Unlog results

NWIC Variable text Full text

Variable Text Size 20

Рис. 39. Анализ частоты отдельных комбинаций слов в рассказе пациента