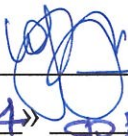


ПРАВИТЕЛЬСТВО МОСКВЫ  
ДЕПАРТАМЕНТ ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ

СОГЛАСОВАНО

Главный внештатный специалист  
по лучевой и инструментальной  
диагностике  
Департамента здравоохранения  
города Москвы

  
Ю. А. Васильев  
«14» февраля 2024 г.

РЕКОМЕНДОВАНО

Экспертным советом по науке  
Департамента здравоохранения  
города Москвы № 3

  
  
«26» февраля 2024 г.

МЕТОДИКА ВАЛИДАЦИИ СРЕДСТВ МЕДИЦИНСКОГО  
АНКЕТИРОВАНИЯ (ОПРОСНИКОВ)

Методические рекомендации № 10

УДК 615.84+616-073.75

ББК 53.6

М 54

Серия «Лучшие практики лучевой и инструментальной диагностики»

Основана в 2017 году

**Организация-разработчик:**

Государственное бюджетное учреждение здравоохранения города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»

**Составители:**

**Васильев Ю. А.** – канд. мед. наук, главный внештатный специалист по лучевой и инструментальной диагностике ДЗМ, директор ГБУЗ «НПКЦ ДиТ ДЗМ»

**Владимирский А. В.** – д-р мед. наук, заместитель директора по научной работе ГБУЗ «НПКЦ ДиТ ДЗМ», профессор кафедры информационных и интернет-технологий ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет)

**Мнацаканян М. Г.** – д-р мед. наук, профессор кафедры госпитальной терапии № 1, заведующая гастроэнтерологическим отделением Университетской клинической больницы № 1 Института клинической медицины им. Н. В. Склифосовского ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет)

**Омелянская О. В.** – руководитель по управлению подразделениями дирекции «Наука» ГБУЗ «НПКЦ ДиТ ДЗМ»

**Решетников Р. В.** – канд. физ.-мат. наук, руководитель отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

**Алымова Ю. А.** – научный сотрудник отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

**Шумская Ю. Ф.** – младший научный сотрудник сектора научных проектов по телемедицине отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

**Ахмедзянова Д. А.** – младший научный сотрудник сектора научных проектов по телемедицине отдела научных медицинских исследований ГБУЗ «НПКЦ ДиТ ДЗМ»

М 54 Методика валидации средств медицинского анкетирования (опросников): методические рекомендации / сост. Ю. А. Васильев, А. В. Владимирский, М. Г. Мнацаканян [и др.] // Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 133. – М.: ГБУЗ «НПКЦ ДиТ ДЗМ», 2024. – 39 с.

**Рецензенты:**

**Панферов Александр Сергеевич** – канд. мед. наук, доцент кафедры факультетской терапии № 1 Института клинической медицины им. Н. В. Склифосовского ФГАОУ ВО Первый МГМУ им. И. М. Сеченова Минздрава России (Сеченовский Университет)

**Гончаров Иван Сергеевич** – канд. мед. наук, заведующий терапевтическим отделением ГБУЗ «ГКБ им. А. К. Ерамишанцева ДЗМ», доцент кафедры внутренних болезней с курсом кардиологии и функциональной диагностики им. В. С. Моисеева Медицинского института ФГАОУ ВО «Российский университет дружбы народов им. Патриса Лумумбы» Минобрнауки РФ.

Методические рекомендации предназначены для научных сотрудников, профессорско-преподавательского состава высших учебных заведений, аспирантов и иных лиц, планирующих и проводящих клинические исследования с использованием опросников.

Данные методические рекомендации разработаны в рамках выполнения научно-исследовательской работы «Научное обоснование моделей и способов организации и оказания медицинской помощи с применением телемедицинских технологий»

*Данный документ является собственностью Департамента здравоохранения города Москвы, не подлежит тиражированию и распространению без соответствующего разрешения*

**ISBN**

© Департамент здравоохранения города  
Москвы, 2024

© ГБУЗ «НПКЦ ДиТ ДЗМ», 2024

© Коллектив авторов, 2024

## СОДЕРЖАНИЕ

НОРМАТИВНЫЕ ССЫЛКИ.....	5
ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ .....	6
ВВЕДЕНИЕ .....	7
ОПРОСНИКИ. ОБЩИЕ СВЕДЕНИЯ .....	8
В ЧЕМ ЗАКЛЮЧАЕТСЯ ПРОЦЕСС ВАЛИДАЦИИ ОПРОСНИКА? .....	14
ЭТАП 1. ЛИНГВОКУЛЬТУРНАЯ АДАПТАЦИЯ.....	17
ЭТАП 2. ПИЛОТНОЕ ИССЛЕДОВАНИЕ .....	19
ЭТАП 3. ВАЛИДАЦИОННОЕ ИССЛЕДОВАНИЕ .....	24
ЗАКЛЮЧЕНИЕ.....	31
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	32
ПРИЛОЖЕНИЕ А .....	35
ПРИЛОЖЕНИЕ Б .....	37
ПРИЛОЖЕНИЕ В.....	38

## **НОРМАТИВНЫЕ ССЫЛКИ**

В настоящем документе использованы ссылки на следующие нормативные документы (стандарты):

1. ГОСТ Р ИСО 14155-2014. Клинические исследования.
2. ГОСТ Р 52379-2005. Надлежащая клиническая практика.
3. ГОСТ Р 56044-2014. Оценка медицинских технологий. Общие положения.

## ТЕРМИНЫ И ОПРЕДЕЛЕНИЯ

В настоящем документе применены следующие термины с соответствующими определениями:

**Опросник** – инструмент, состоящий из набора вопросов и/или утверждений, который используется для сбора различной информации об индивидах, его заполняющих.

**Пациент** – физическое лицо, которое обратилось за оказанием медицинской помощи и которому оказывается медицинская помощь, в том числе с использованием телемедицинских технологий.

**Респондент** – лицо, принимающее участие в опросе, анкетировании.

**Надежность методики** – устойчивость тестовых показателей относительно случайных факторов или особенностей конкретного пациента.

**Достоверность методики** – показатель, который отображает меру соответствия результатов опросника поставленным задачам.

**Телемедицинские технологии** – информационные технологии, обеспечивающие дистанционное взаимодействие медицинских работников между собой, с пациентами и (или) их законными представителями, идентификацию и аутентификацию указанных лиц, документирование совершаемых ими действий при проведении консилиумов, консультаций, дистанционного медицинского наблюдения за состоянием здоровья пациента.

## ВВЕДЕНИЕ

Опросник – это универсальный инструмент, который широко используется для получения необходимых данных от респондентов как в ходе научных исследований, так и в клинической практике. Опросники нужны для сбора соответствующей информации с целью измерения каждого из элементов области интереса, а также для решения задач исследования. Использование опросников может быть сопряжено с социокультурными особенностями конкретной популяции, что требует предварительной адаптации зарубежных опросников. Если оригинал анкеты на английском должен быть передан группе респондентов, не владеющих этим языком, необходимо обеспечить первоначальный перевод и последующую валидацию полученной версии опросника. Валидация иноязычного опросника – это научный процесс, включающий (1) перевод вопросов/пунктов с иностранного языка на предназначенный для местного населения и (2) последующую демонстрацию того, что переведенная версия обладает достаточным уровнем надежности и достоверности, путем ее применения к репрезентативной выборке местного населения.

Возможны сценарии, в которых содержание оригинальной версии опросника на иностранном языке не устраивает представителей местного населения, поскольку многие факторы потенциально могут способствовать различному восприятию или пониманию контекста. Различия в понимании некоторых концептов могут быть вызваны расхождениями в индивидуальных религиозных убеждениях, культурном наследии, образе жизни, социальной и бытовой среде и т. д. Таким образом, при переводе опросника на другой язык можно получить нерелевантные ответы. Это означает, что проведение исследования с использованием опросника, который был просто переведен на другой язык, но не прошел надлежащую валидацию, является нецелесообразным. К сожалению, основная масса опросников, используемых в научных исследованиях в России, не проходили процесс валидации на русскоговорящем населении.

Основной целью данных методических рекомендаций является освещение методики валидации уже разработанных иностранных опросников.

## ОПРОСНИКИ. ОБЩИЕ СВЕДЕНИЯ

Опросник – это инструмент, состоящий из набора вопросов и/или утверждений, который используется для анкетирования – сбора различной информации об индивидах, его заполняющих (например, их мнение, поведение, физиологические показатели и др.) [1].

Анкетирование широко применяется не только в медицинских исследованиях, но также и в клинической медицине. Опросники могут использоваться в качестве единственного исследовательского инструмента, а могут дополнять иные методы.

Существует 4 основных типа опросников по области применения:

### **1. Опросники 1-го типа: для сбора информации.**

Опросники первого типа разработаны так, чтобы получать от респондентов объективную информацию и факты. Каждый вопрос является независимым и не связан с другими. При помощи них можно получить информацию о социально-демографических характеристиках пациента, таких как пол, этническая принадлежность и данные анамнеза (сопутствующие заболевания, история курения, результаты лечения) и т. д. Полученная информация может быть использована как для эпидемиологических исследований, так и для клинических в совокупности с другими данными [1].

### **2. Опросники 2-го типа: для оценки уровня знаний.**

Второй тип опросников создан для проверки знаний респондентов. Примером является экзаменационный тест, который содержит набор узкоспециализированных вопросов, относящихся к конкретной теме. Чаще всего они начинаются со слов «что», «кто», «когда», «почему» и «как». Для вычисления общего результата, отражающего уровень знаний респондентов, опросники имеют специальную формулу подсчета ответов. Опросники второго типа, как правило, используются с целью ранжирования респондентов [1].

### **3. Опросники 3-го типа: для скрининга и диагностической оценки.**

Этот тип опросника связан с конкретными фактами и симптомами и направлен на выявление связи между зависимыми и независимыми переменными (исходом и факторами, которые потенциально способны на него повлиять). Ответы, полученные от респондентов, преобразуются в баллы. Заранее определенный граничный балл позволяет установить критерии наличия заболевания. Такие опросники часто используются в клинической медицине. С их помощью можно заподозрить наличие заболевания, например депрессивного состояния, в психиатрии. При других патологиях для подтверждения диагноза требуется проведение



лабораторно-инструментальных методов обследования, в таком случае допустимо проводить скрининг с помощью опросника с целью выявления целевой выборки. Для этого, помимо исследования достоверности и надежности адаптированной версии опросника, необходимо также подтвердить и его диагностическую/прогностическую ценность [2].

#### **4. Опросники 4-го типа: для измерения «скрытых» переменных.**

Четвертый тип опросников используется для измерения «скрытых» переменных, которые не могут быть измерены с помощью объективных методов. К ним относятся: качество жизни, уровень тревоги и депрессии, уровень удовлетворенности различными аспектами жизни. Для оценки таких переменных следует использовать как минимум два вопроса, хотя в некоторых специализированных исследованиях используются 10 и более вопросов [3, 4]. Для измерения конкретной «скрытой» переменной посредством опросника исследователю необходимо с помощью инструмента исследования получить значение «скрытой» переменной из наблюдаемой переменной. Хорошо известно, что опросники, измеряющие «скрытые» переменные, являются наиболее сложными для разработки и валидации. В отличие от первого и второго типов опросников, корректное измерение «скрытых» переменных является нетривиальной задачей, так как они интерпретируются субъективно [1].

Выделяют следующие виды опросников по структуре:

- Открытый – позволяет респонденту отвечать на вопрос своими словами. Вопросы обычно начинаются со слов: «как», «почему» и «что» и побуждают респондента давать более подробные и персонализированные ответы.
- Закрытый – предоставляет респонденту при выборе ответа набор заранее определенных вариантов. Опросники такого типа легче анализировать и обобщать.
- Смешанный – это комбинация из открытых и закрытых вопросов.
- Иллюстрированный – вопросы представлены в виде схем, диаграмм, изображений и др.
- Структурированный – имеет заранее определенные, обычно закрытые, вопросы.
- Неструктурированный – не имеет фиксированного формата или заранее определенных вопросов. Вместо этого интервьюер может задавать респонденту открытые вопросы и позволять ему отвечать в свободной форме.

Также могут отличаться типы вопросов в опроснике:

- Вопросы с несколькими вариантами ответов.
- Вопросы с оценкой ответов по шкале от 1 до 10.
- Открытые вопросы.
- Вопросы с оценкой ответов по шкале Ликерта («совершенно согласен»; «согласен»; «затрудняюсь ответить / нейтральное отношение»; «не согласен»; «совершенно не согласен»).
- Демографические вопросы (касаются личной информации участника, например, возраст, пол, этническая принадлежность, уровень образования и т. д.).
- Вопросы с вариантами ответов «да/нет».
- Рейтинговые вопросы, в которых необходимо ранжировать несколько элементов в порядке предпочтения или важности.
- Матричные вопросы, представляющие собой матрицу или сетку вариантов, из которых респонденты могут выбрать наиболее подходящий ответ.
- Дихотомические вопросы, представляющие два варианта, которые являются противоположными или противоречащими друг другу. Они полезны для измерения бинарных или поляризованных установок.

Для того чтобы использовать в своем исследовании валидный опросник, исследователь может пойти двумя путями: разработать собственный или валидировать уже существующий. Разработка опросника – сложный процесс, состоящий из нескольких этапов, которые включают в себя оценку надежности и валидности (достоверности). Понятия «валидность» и «достоверность» равнозначны, однако в дальнейшем для избежания путаницы в тексте данных методических рекомендаций мы будем использовать только термин «достоверность».

## **ВАЖНО!**

*Во время работы с валидацией опросников нужно понимать, что входит в такие понятия, как «надежность» и «достоверность» инструмента. Надежность методики – это устойчивость тестовых показателей относительно случайных факторов или особенностей конкретного пациента. Достоверность – это показатель, который отображает меру соответствия результатов опросника поставленным задачам.*

*Например, весы, которые каждый раз показывают один и тот же результат при взвешивании одного и того же человека (то есть демонстрируют высокую надежность), необязательно отражают*

истинную (достоверную) массу человека. Поэтому для подтверждения того, что весы дадут точное измерение массы (доказательства их достоверности), исследователю необходимо сравнить их с показателями «золотого стандарта» (т. е. ранее откалиброванных весов).

Не всегда (и даже – довольно редко) для оценки интересующего показателя требуется создание собственного опросника. Существует резерв опросников для оценки определенных параметров, обладающих проверенной надежностью и достоверностью. Поскольку большинство из них являются англоязычными, требуется их предварительная лингвокультурная адаптация на русский язык и проведение валидационного исследования. При этом сложность валидации уже разработанных иными исследователями и рабочими группами опросников зависит от их типа (рисунок 1).

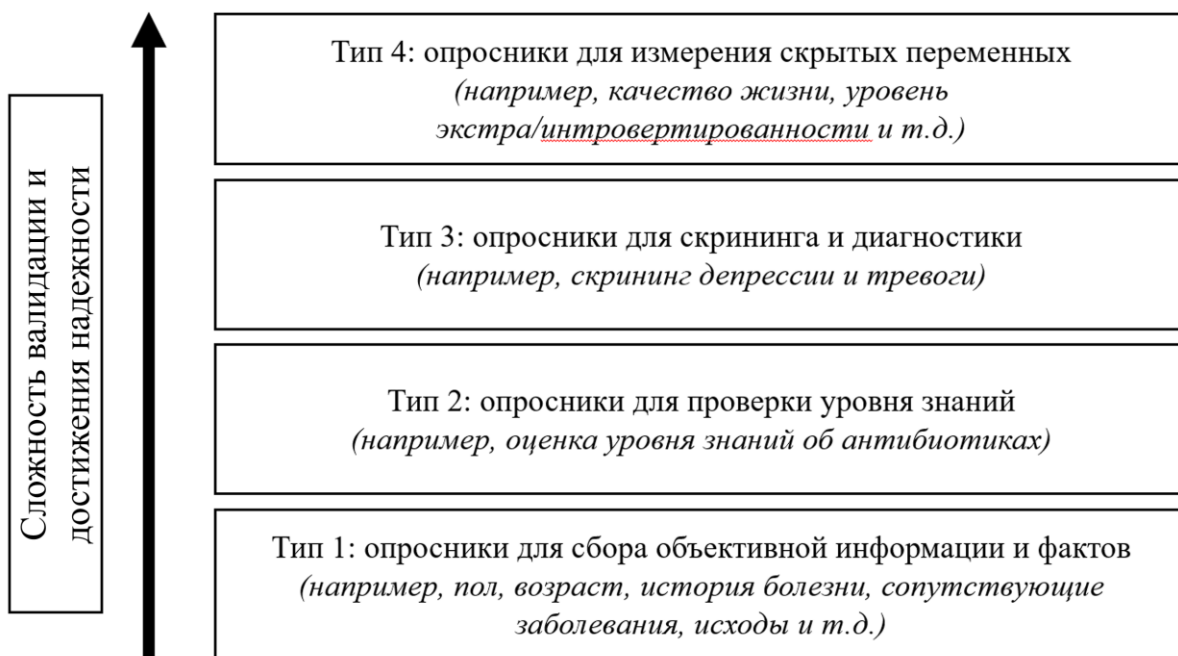


Рисунок 1 – Оценка сложности валидации различных типов опросников

### **В каких случаях отсутствует необходимость проводить валидацию опросников?**

Иногда возникают ситуации, когда перевод и последующая валидация переведенного опросника не нужны. Это возможно в том случае, если соблюдаются следующие условия:

1. У популяции в первичном исследовании опросника (при его создании) такой же уровень понимания русского языка, как и у русскоязычной популяции.

2. Все вопросы соответствуют их знаниям, компетенциям, личному опыту и ситуации.

3. Ни один из вопросов не подразумевает ответов, которые могут отклоняться от нормы из-за различий:

- а) в используемой терминологии;
- б) культурных особенностях;
- в) образе жизни;
- г) жизненной и социальной среде;
- д) религиозных убеждениях.

***Пример:** Опросник образа собственного тела (ООСТ), направленный на оценку удовлетворенности собственным телом у лиц, страдающих расстройствами пищевого поведения. Данный инструмент был разработан О. А. Скугаревским и С. В. Сивухой в Белорусском государственном университете. Опросник был создан на русском языке и валидирован на белорусской популяции. В Беларуси русский язык является государственным, а уровень его понимания населением очень высокий. Таким образом, данный опросник можно использовать и на российской популяции.*

### **Типы опросников, которые могут не подойти для валидации**

Не все опросники пригодны к валидации для группы пациентов, говорящих на другом языке. Некоторые из этих опросников изначально были опубликованы без предоставления достаточных доказательств их достоверности и надежности. Как правило, исследователь создает опросник, а затем публикует его одновременно с информацией о разработке. Такая работа может быть представлена в виде научной статьи, книги или отчета. Соответственно, не рекомендуется проводить валидационное исследование для опросников, не имеющих четких данных о процедуре разработки.

***Пример:** Опросники, созданные под конкретное исследование, без отдельной работы по оценке достоверности самого опросника.*

### **Одновременное проведение валидационного и клинического исследований**

Существует практика совмещения валидационного и клинического исследований в одном. В этой ситуации в клиническом исследовании

используется еще не валидированный, но переведенный иностранный опросник, который проходит процесс валидации во время клинического исследования. После завершения исследования публикуются результаты как такой упрощенной валидации, так и самого клинического исследования. Такая практика имеет свои недостатки, так как многие пункты полноценной валидации опросника остаются невыполненными. Объединение клинического и валидационного исследований возможно в случае, если критерии включения и исключения в них совпадают [1].

***Пример:** Валидация нового опросника для оценки качества жизни пациентов с онкологическим заболеванием может проводиться в исследовании оценки качества жизни этих пациентов. В таком случае следует использовать как минимум два опросника: валидируемый одновременно с уже доказавшим свою достоверность и надежность.*

## **В ЧЕМ ЗАКЛЮЧАЕТСЯ ПРОЦЕСС ВАЛИДАЦИИ ОПРОСНИКА?**

Первым шагом в проведении валидационного исследования является выбор опросника. Существует большое число опросников практически для каждого параметра. К примеру, только для оценки приверженности лекарственной терапии известно более десяти, для некоторых из них при этом недоступны данные первичного валидирования, для других – нет четкой информации о надежности и достоверности. При этом может быть и обратная ситуация, когда подходящих опросников несколько. Какому отдать предпочтение? Об этом будет рассказано далее.

Сам же процесс валидации, как уже было сказано, состоит из нескольких обязательных этапов:

- первый этап – лингвокультурная адаптация;
- второй этап – пилотное исследование;
- третий этап – валидационное исследование (исследование достоверности).

Схематически процесс представлен на рисунке 2.

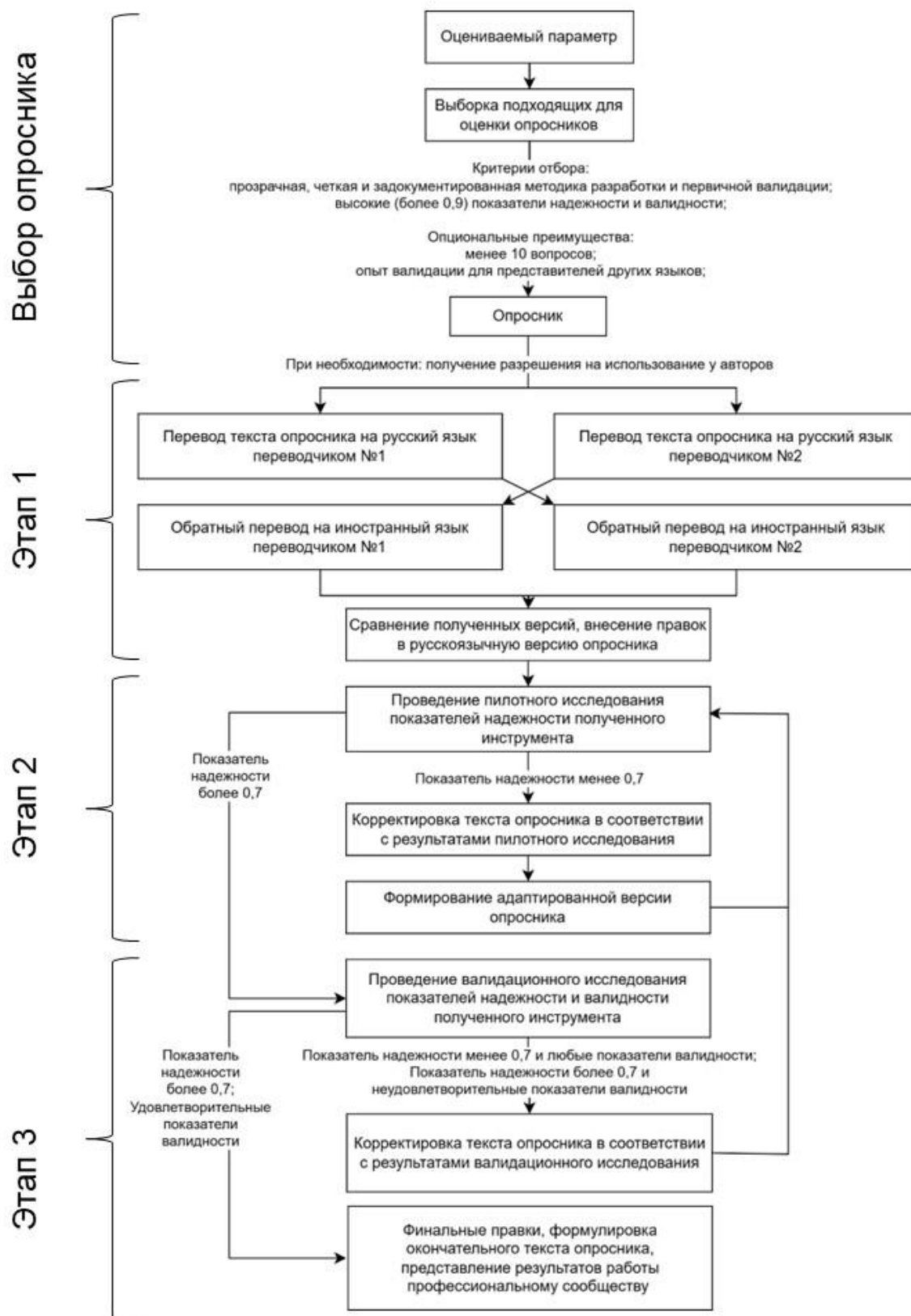


Рисунок 2 – Схема выполнения валидации иноязычного опросника

Каждый этап валидации мы рассмотрим в следующих главах.

**ВАЖНО!**

*Если после проведения первоначальной оценки достоверности содержания опросника выясняется, что в нем слишком много пунктов, требующих модификации, то исследователям рекомендуется выбрать другой подходящий опросник, если таковой имеется. В противном случае единственным приемлемым вариантом для исследователя может стать разработка нового опросника.*



## ЭТАП 1. ЛИНГВОКУЛЬТУРНАЯ АДАПТАЦИЯ

**Цель:** получить адаптированную к русскоязычному респонденту версию опросника без потери смысла.

**Методика:** существует множество подходов к переводу опросников. Одним из лучших является привлечение двух профессиональных переводчиков, работающих независимо друг от друга для выполнения прямого, а затем обратного перевода [2, 5–7] (см. рисунок 2).

После выбора подходящего иностранного опросника необходимо произвести его перевод с адаптацией (при необходимости) терминов и понятий двумя переводчиками независимо друг от друга. При этом рекомендуется, чтобы хотя бы один из переводчиков имел опыт перевода в сфере профессиональной коммуникации или являлся экспертом в целевой области [6, 8]. Эксперты в определенной области могут излишне использовать научные термины, затрудняя понимание опросника обывателями. В то же время лингвисты, в связи с недостаточной медицинской осведомленностью, могут исказить суть опросников.

Вторым этапом следует провести обратный перевод полученных версий на иностранный язык (при этом должен произойти обмен переведенными версиями, чтобы Переводчик № 2 переводил версию Переводчика № 1, и наоборот). Полученные в результате четыре версии опросника нужно сопоставить между собой и оценить. Для этой цели создается экспертная группа, в состав которой войдут эксперты в области использования опросника, а также профессиональный медицинский переводчик, до этого не участвовавший в процессе перевода. Итогом работы экспертной группы станет адаптированная консенсусная версия исходного иноязычного опросника. Критерий принятия каждого вопроса и общего опросника – согласие между экспертами выше 75 % (то есть каждый элемент должен набрать как минимум 75 голосов, если в консенсусе участвуют 100 экспертов).

**Пример:** В опроснике *Personal Wellbeing Index – Adult (PWI-A)* для оценки индекса личного благополучия взрослых присутствует вопрос «*How satisfied are you with your spirituality or religion?*», дословный перевод которого звучит так: «*Насколько Вы удовлетворены своей духовностью или религиозностью?*». Для русскоязычной популяции термин «духовность» имеет светское и религиозное значение. В первом случае оно сводится к моральным качествам, таким как нравственность, добродетельность и интеллектуальность, а во втором – отражает религиозность человека, его практику воплощения личного акта веры.

После успешного завершения лингвокультурной адаптации следующим этапом является пилотное исследование.

### **ВАЖНО!**

*В литературе имеются упоминания о прецедентах, когда перевод опросника осуществлялся одним переводчиком – экспертом в целевой области. В этом случае переводчик должен знать иностранный язык на уровне носителя. Такой подход наиболее приемлем для простых опросников, целью которых является сбор фактической информации (опросники 1-го и 2-го типа) [1]. Но проведение прямого и обратного перевода при валидации опросников в настоящее время является общепринятой нормой.*

## ЭТАП 2. ПИЛОТНОЕ ИССЛЕДОВАНИЕ

**Цель:** проверить надежность (для всех типов опросников) и внутреннюю согласованность (для опросников типов 3 и 4) адаптированной версии опросника на небольшой группе респондентов.

**Выборка:** выборка в пилотном исследовании должна обладать теми же характеристиками, что и в валидационном исследовании, то есть быть репрезентативной. Предположим, что целью исследования является оценка качества жизни пациентов с сахарным диабетом 2-го типа. В пилотном исследовании среди не менее 30 респондентов [1] выборка должна иметь равное распределение пациентов по характеристикам заболевания, таким как: (i) недавно диагностированный сахарный диабет 2-го типа, (ii) длительно существующий сахарный диабет 2-го типа, (iii) плохо контролируемый сахарный диабет 2-го типа, (iv) сахарный диабет 2-го типа с осложнениями и т. д. Ответы респондентов, обладающих широким спектром сильно различающихся характеристик, позволят обеспечить объективность и точность полученных данных. Нельзя собирать данные только от здоровых респондентов, так как в этом случае итоги опросника не будут отображать результаты, обусловленные различиями в состоянии здоровья респондентов.

Формирование «удобной» или «стихийной» выборки (convenience sampling – подход, при котором исследователь набирает добровольцев из группы людей, соответствующих общим требованиям проведения исследования) обычно является предпочтительным для использования в пилотном исследовании.

### **Методика:**

#### **1. Надежность.**

При работе с опросниками понятие «надежность» подразумевает ответ на два вопроса [9]:

А. Насколько хорошо согласуются между собой повторные измерения – насколько они воспроизводимы?

В. Насколько опросник позволяет отличить респондентов друг от друга?

Для валидационных исследований иноязычных опросников под проверкой надежности понимают оценку ретестовой воспроизводимости (test-retest reliability). Основная идея состоит в том, чтобы проверить, насколько похожи будут полученные в двух разных временных точках ответы одного и того же респондента на один и тот же вопрос. Хорошие показатели надежности достигаются при одновременном соблюдении всех условий:

- Вопрос однозначно понимается респондентом, при повторном тестировании вероятность понимания вопроса иным образом минимальна.
- Факторы, которые могут повлиять на ответы респондента, оставались в период между измерениями постоянными (социальное благополучие, условия жизни, состояние здоровья и т. д.).
- Вопрос не влияет на состояние и мнение респондента об исследуемом предмете (формулировка вопроса не заставляет респондента изменить свое мнение или впечатление, что могло бы отразиться на ответе на такой же вопрос при повторном тестировании).

### **ВАЖНО!**

*Еще на этапе планирования исследования по валидации опросника целесообразно оценить влияние условий 2 и 3 – и минимизировать их, в таком случае получившиеся значения ретестовой надежности можно трактовать однозначно: насколько хорошо вопросы понимаются респондентами.*

Не существует единого мнения об оптимальной длительности перерыва между первичным и повторным тестированием, он должен определяться индивидуально для каждого опросника. Необходимо, с одной стороны, снизить вероятность изменений в жизни респондента (стабильность оценки), а с другой – обеспечить достаточный перерыв, чтобы ответы на первое тестирование были забыты и не влияли на выбор при повторном тесте (независимость испытаний). В большинстве случаев оптимальный срок – 7–14 дней, однако он может меняться при наличии оснований и достаточной обоснованности [5, 10].

Выбор статистических методов для оценки ретестовой надежности зависит от типа вариантов ответа на вопросы опросника.

1) Ответ в виде непрерывных значений. *Пример: ответ в свободной форме на вопрос «Сколько часов Вы фактически отработали в течение последних семи дней?».* Общепринятым подходом в случае ответов такого типа является определение коэффициента внутриклассовой корреляции (Interclass Correlation Coefficient, ICC). Существует несколько форм расчета ICC, зависящих от условий тестирования (Приложение 2). Значение ICC находится в интервале между 0 и 1, где 1 означает полную ретестовую надежность, 0 – отсутствие воспроизводимости и низкую ретестовую надежность [11].

Граничные значения коэффициента внутриклассовой корреляции:

- <0,5 – слабая (poor) надежность;
- 0,5–0,75 – умеренная (moderate) надежность;

- 0,75–0,9 – хорошая (good) надежность.

2) Ответ в виде категориальных значений. *Пример: «Как Вы себя чувствуете?» – хорошо/удовлетворительно/плохо; «Был ли у Вас инсульт в анамнезе?» – да/нет.* Для оценки согласованности ответов, выраженных категориальным образом, чаще всего используется коэффициент каппа Коэна (Приложение 3). [12]. В зависимости от того, к какому типу относятся категориальные данные, необходим расчет этого коэффициента в разных модификациях:

- Стандартная каппа Коэна: категориальные данные номинального типа (равнозначные, взаимоисключающие, независимые варианты). *Примеры: «да/нет», «болен/здоров».*

- Взвешенная каппа Коэна: порядковые категориальные данные (ответы можно расположить в относительном друг от друга порядке). *Примеры: «часто/редко/никогда», шкала от 1 до 9.*

Использование той или иной каппы Коэна позволяет учесть отличие между категориями ответов в зависимости от того, насколько «далеко» категории расположены друг от друга. Так, в случае примера «часто/редко/никогда», разница между ответами «часто» и «редко» будет менее существенной, чем между «часто» и «никогда» – в таком случае каждой разнице присваиваются свои «веса», численно отражающие ее значимость. Стандартом является использование квадратичных весов, однако этот вопрос остается на усмотрение исследователя. Существуют также модификации взвешенной каппы Коэна с линейными и другими типами весов [13].

Удовлетворительным показателем надежности принята каппа Коэна более 0,8 [14].

Граничные значения каппы Коэна:

- 0–0,2 – слабое (weak) согласие;
- 0,21–0,4 – незначительное (fair) согласие;
- 0,41–0,6 – умеренное (moderate) согласие;
- 0,61–0,8 – высокое (substantial) согласие;
- 0,81–1 – полное (almost perfect) согласие.

Статья об оригинальном опроснике часто содержит информацию о том, какой метод статистической обработки использовался при его создании. Исследователю рекомендуется использовать тот же метод для адаптированной версии опросника, но при необходимости он может прибегнуть к дополнительному использованию и других методов оценки надежности (порядковая альфа [15, 16], омега Макдональда [17],  $r$  Райкова [18, 19] или бета Ревелла [20, 21], оценка по методу split-half, формула

Спирмена-Брауна, метод альтернативных форм (коэффициент эквивалентности), межнаблюдательная надежность [18, 19]).

Граничные значения коэффициентов, используемых для проверки надежности опросников, эмпирические и, как правило, ничем не обоснованы [22]. В тех прикладных ситуациях, когда по результатам тестирования принимаются важные решения, надежность 0,90 является минимально допустимой, а надежность 0,95 следует считать рекомендуемым стандартом [23, 24]. Для групповых оценок и исследовательских целей достаточной считается надежность выше 0,70 [23, 24].

**2. Внутренняя согласованность (*Internal consistency*)** также является мерой, определяющей общую надежность опросника наряду с ретестовой надежностью. Она определяется связью каждого конкретного вопроса или домена опросника с общим результатом, тем, насколько каждый элемент анкеты входит в противоречие с остальными, насколько каждый отдельный вопрос измеряет признак, на оценку которого направлен весь опросник.

Для оценки внутренней согласованности наиболее распространенным методом является альфа Кронбаха, однако в настоящее время начинают получать признание и другие показатели:  $r$  Райкова [19], порядковая альфа и бета Ревелла [20, 21].

Коэффициент альфа Кронбаха можно использовать для опросников, имеющих как минимум один домен (Приложение 4). Это хороший метод для опросников, измеряющих скрытые переменные [7], однако применение этого метода ограничено требованием к нормальности распределения данных. Интерпретация значения альфы Кронбаха следующая:

- $\leq 0,5$  – недостаточная (unacceptable) надежность;
- $> 0,5$  – слабая (poor) надежность;
- $> 0,6$  – сомнительная (questionable) надежность;
- $> 0,7$  – достаточная (acceptable) надежность;
- $> 0,8$  – высокая (good) надежность;
- $> 0,9$  – очень высокая (excellent) надежность.

### **ВАЖНО!**

*Если в пилотном исследовании выявлено, что надежность опросника низкая, то исследователю необходимо еще раз тщательно проанализировать все вопросы и выявить истинную причину низкой надежности (т. е. связано ли это с неточным переводом оригинального опросника, или с неадекватной адаптацией фактического содержания оригинального опросника к конкретной группе местных респондентов).*

*После этого текст необходимо переработать и **выполнить пилотное исследование еще раз.***

В зависимости от результатов, полученных в ходе пилотного исследования, исследователь может:

- изменить содержание опросника до проведения валидационного исследования;
- в случае удовлетворительных результатов пилотного исследования перейти к следующему этапу – непосредственно к валидационному исследованию с использованием большой выборки респондентов.

### ЭТАП 3. ВАЛИДАЦИОННОЕ ИССЛЕДОВАНИЕ

**Цель:** определение надежности, а также достоверности адаптированной версии опросника на большой выборке респондентов.

**Выборка:** в валидационном исследовании число участников строго зависит от типа опросника и, соответственно, от того, какие статистические методики будут применяться.

**Методика:** в ходе валидационного исследования следует придерживаться той же методологии, что и во время создания валидируемого опросника. Исследователю нужно ознакомиться со статьей или отчетом о ходе создания оригинального опросника для выяснения следующих деталей:

- 1) критерии включения и невключения в исследование;
- 2) дизайн исследования и методология набора в исследование;
- 3) метод статистического анализа.

После уточнения указанной выше информации нужно решить, повторять ли дизайн исходного исследования, или скорректировать его. Различия в дизайне между оригинальной работой и валидационным исследованием переведенной версии опросника могут привести к различиям в результатах. С другой стороны, часть опросников была разработана достаточно давно, за прошедшее с этого момента время как наука, так и социум развивались и изменялись, что может требовать переработки уже существующего инструмента.

Надежность в валидационном исследовании определяется тем же образом, что и в пилотном исследовании. Но в отличие от пилотного исследования, в валидационном также определяется достоверность опросника. Виды определяемой достоверности зависят от типа опросника (таблица 1).



Таблица 1 – Метрики, которые рекомендованы к оценке в ходе валидационного исследования различных типов опросников

Методы подтверждения надежности и достоверности	Тип опросника			
	Сбор фактов и объективной информации (тип 1)	Измерение уровня знаний (тип 2)	Измерение «скрытых» переменных (тип 3)	Скрининг или диагностика (тип 4)
Надежность	+	+	+	+
Внутренняя согласованность			+	+
Содержательная достоверность	+	+	+	+
Очевидная достоверность	+	+	+	+
Достоверность по известным группам		+	+	+
Параллельная достоверность		+	+	+
Прогностическая достоверность		+	+	+
Конструктивная достоверность			+	
Диагностическая оценка				+

### **ВАЖНО!**

*Так же как и в пилотном исследовании, при низкой надежности и/или достоверности опросников исследователю следует переформулировать пункты анкеты.*

Далее мы раскроем подробнее указанные виды достоверности.

1. **Содержательная достоверность (content validity)** – это оценка способности каждого элемента/вопроса оценить то, что он должен измерять. Если опросник содержит вопросы, связанные с восприятием, чувствами, культурными убеждениями, выбором образа жизни или любыми другими аспектами человеческого поведения, то обязательно нужно определять содержательную достоверность переведенной версии.

Эта операция может проводиться экспертом (профессионалом, работающим в модальности, оценке которой посвящен опросник) или группой экспертов коллегиально [25].

*Пример:* Если опросник связан с качеством жизни пациентов с сахарным диабетом, то подходящим экспертом для определения его достоверности будет эндокринолог или врач общей практики.

При повторной оценке достоверности содержания переведенного варианта анкеты исследователь может принять одно из следующих решений:

- Все пункты оригинальной анкеты релевантны, и оригинальный вариант анкеты сохраняется. В таком случае первоначальный вариант анкеты полностью совпадает с ее переведенной версией [2, 7].

- Большинство пунктов релевантны, за исключением нескольких, которые следует перефразировать. То есть содержание переведенного варианта анкеты будет несколько отличаться от оригинальной версии [6].

- Добавление и/или исключение одного или нескольких вопросов. Это означает, что переведенная версия анкеты будет отличаться от оригинальной версии [26].

Сохранение первоначального варианта опросника помогает сохранить оригинальный метод подсчета баллов (если такой имеется), чтобы результаты опроса были сопоставимы. Но если это не представляется возможным, то можно модифицировать исходный вариант опросника [5].

2. **Очевидная достоверность (*face validity*)** – степень, в которой опросник воспринимается потенциальными респондентами как адекватный для понимания [27]. Для ее оценки необходима обратная связь от респондентов, полученная с помощью интервьюирования. Опросник выдается респондентам, после этого с ними проводится личная беседа о масштабах исследования, методике, формате ответов и всех пунктах опросника. В ходе беседы участникам необходимо задать следующие вопросы:

- Понятно ли, по их мнению, что имеется в виду в вопросе?
- Могут ли они переформулировать его своими словами?
- Как они пришли к своему ответу?
- Что они связывают с конкретными словами, используемыми в формулировке?

- Есть ли в вопросе слова, которые они не поняли или сочли неприемлемыми?

В случае наличия альтернативных вариантов формулировок респонденты могут высказать свои предложения, чтобы они точно воспринимали вопросы так, как это планировалось создателями оригинальной версии опросника.

Полученные результаты обобщаются в итоговый отчет, включающий соответствующие сведения об участниках, а также заключительные рекомендации по совершенствованию инструмента. Этот процесс гарантирует, что переведенная версия оригинального опросника будет максимально отражать его изначальный смысл.

### **ВАЖНО!**

*Обычно в процессе оценки очевидной достоверности участвуют около 10 респондентов, чтобы можно было реализовать индивидуальный подход к каждому.*

3. *Достоверность по известным группам (differentiation by known groups)* демонстрируется в тех случаях, когда опросник может различать две группы, которые отличаются по интересующей переменной [28]. Сравнение известных групп может проводиться для валидации отдельного вопроса или домена. Вопрос или домен считается валидным, если две популяции с определенными и четко различающимися характеристиками имеют статистически значимые различия в среднем или медианном балле.

*Пример: Опросник Кесслера для разграничения людей с нарушением психологической адаптации и с нормальной адаптацией.*

4. *Параллельная достоверность (concurrent validity)* – это мера достоверности, при которой другой, ранее валидированный инструмент используется в качестве ориентира для оценки достоверности нового [9]. Для того чтобы оценить параллельную достоверность нового опросника, необходимо, чтобы был валидирован и опубликован созданный ранее и широко используемый для этой цели опросник. Этот валидированный опросник должен служить базой для валидации нового. Общепринятым статистическим методом проверки параллельной достоверности является определение корреляции между одними и теми же доменами из двух разных опросников.

*Пример:* Широко известный опросник SF-36 (оценка качества жизни) использовался для валидации *Diabetes Quality of Life* (оценка качества жизни пациентов с сахарным диабетом) [30].

5. **Прогностическая достоверность (predictive validity)** – это мера достоверности, которая заключается в определении того, насколько хорошо измерение одной переменной может предсказать значение другой переменной. Тестовый балл, полученный по опроснику, взят в момент времени 1. Для определения прогностической достоверности этого опросника нужно проверить, сможет ли балл, полученный в момент времени 1, предсказать другое событие в будущем. Поскольку часто для применения этого метода необходим долгий период наблюдения, он не является широко распространенным. Тем не менее прогностическая достоверность остается весьма полезной мерой, позволяющей оценить, насколько вероятно, что результаты теста, указанные в опроснике, будут предсказывать определенные события в будущем.

Для определения прогностической достоверности также используется корреляционный анализ взаимосвязи между значениями двух исследуемых переменных.

*Пример:* Высокий балл в опроснике оценки индекса висцеральной чувствительности может привести к низкому значению показателей опросника качества жизни в будущем. В зависимости от того, насколько сильно будут коррелировать результаты индекса с результатами качества жизни, оцененными через 6–12 месяцев, можно будет говорить о прогностической достоверности опросника оценки индекса висцеральной чувствительности.

6. **Конструктивная достоверность (construct validity)** часто рассматривается как один из предпочтительных методов определения достоверности «скрытых» (неподвластных объективному измерению аппаратом или устройством) переменных, измеряемых некоторыми опросниками. Примером таких переменных служат качество жизни, комплаентность – они не могут быть объективно измерены с помощью какой-либо меры. Наилучшим способом измерения этих переменных является использование надежного и достоверного опросника.

Конструктивная достоверность определяет степень, в которой баллы, полученные по валидируемому тесту, коррелируют с баллами, полученными по другим тестам, предназначенным для оценки того же параметра. В то же время она призвана показать, в какой степени

показатели опросника, которые теоретически не имеют высокой связи, и в самом деле оказываются не связанными друг с другом [1].

Для опросников, измеряющих скрытые переменные, конструктивная достоверность должна быть подтверждена факторным анализом. Детальная методология проведения факторного анализа выходит за рамки настоящих методических рекомендаций; с нею можно ознакомиться, например, в материале Янга с соавторами [31]. Суть факторного анализа схематично представлена на рисунке 3: в данном примере восемнадцать вопросов используются для оценки трех различных скрытых переменных (допустим, физического, психического и социального благополучия). Хороший инструмент состоит из вопросов, которые оценивают именно свою область. Это позволяет точно определить общие факторы, которые могут объяснить порядок или структуру между измеряемыми переменными. Например, согласно рисунку 3, восемь вопросов, окрашенных в зеленый цвет, соответствуют фактору 1 (физическое благополучие), пять вопросов, окрашенных в фиолетовый цвет, – фактору 2 (психическое благополучие), а шесть вопросов, окрашенных в красный цвет, – фактору 3 (социальное благополучие). В сомнительном инструменте вопросы смешаны и могут относиться к оценке разных факторов.



Рисунок 3 – Работа факторного анализа. Каждый вопрос, относящийся к оценке той или иной переменной, отмечен соответствующим цветом

Концепция конструктивной достоверности заключается в том, что если группа пунктов опросника измеряет один показатель, то эти пункты могут быть объединены в единый домен. Кроме того, если все пункты опросника измеряют более одного показателя, то предлагаемый статистический анализ должен быть способен различить эти пункты и разделить их на разные домены.

*Пример:* Краткая версия опросника DASS состоит из трех основных доменов: депрессия, тревога и стресс, с семью пунктами для каждого домена [32]. Поэтому любые валидационные исследования, которые будут проводиться для определения конструктивной достоверности данного опросника, переведенного на все другие языки, также должны быть направлены на выявление трех основных доменов, причем каждый домен в идеальном варианте должен содержать такое же количество пунктов, как и в оригинальном опроснике [33, 34].

Проблема, с которой обычно сталкиваются при определении конструктивной достоверности опросника, заключается в том, что показатели, полученные в результате адаптации опросника, не совпадают с теми, которые были заложены в исходный опросник. Если предположить, что дизайн и методология исследования (такие как социально-демографические характеристики исследуемой совокупности, общий процесс перевода, выборочная совокупность, выбор методов выборки и размер выборки) полностью совпадают с теми, которые использовались при разработке исходного опросника, то такие различия могут быть обусловлены расхождениями в восприятии контекста разными респондентами, то есть это может быть связано с лингвокультурными особенностями респондентов. Если такая проблема возникает, то необходимо минимизировать различия в смысловом содержании опросника.

Таким образом, после успешного выполнения трех этапов валидации исследователь получает русскоязычный опросник с высокими показателями надежности и достоверности.

## ЗАКЛЮЧЕНИЕ

Опросники – это уникальный диагностический и исследовательский инструмент, однако его использование требует соблюдения определенных правил подготовки. Для применения в валидационном исследовании подходят только те опросники, которые уже были опубликованы и валидированы ранее. В связи с этим исследователь должен иметь представление о процессе разработки и конструирования оригинального опросника. При валидации переведенной версии опросника исследователю придется выполнять практически все те же процедуры, что и при его разработке, поэтому очень важно с самого начала убедиться в правильности выбранной методологии исследования (для разработки и валидации опросника), включая выбор исследовательского дизайна и предлагаемого плана статистического анализа. Однако нельзя не отметить тот факт, что исследователь всегда имеет возможность при необходимости представить любые дополнительные доказательства надежности и достоверности опросника, чтобы еще убедительнее подтвердить его обоснованность и универсальность.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Bujang M. A., Khee H. Y., Yee L. K. A Step-By-Step Guide to Questionnaire Validation Research. Institute for Clinical Research, NIH, Malaysia, 2022.
2. Yunus A. et al. Validation of the Malay version of Berlin questionnaire to identify Malaysian patients for obstructive sleep apnea // Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia. 2013. Vol. 8, № 1. P. 5.
3. DCCT Research Group. Reliability and validity of a diabetes quality-of-life measure for the diabetes control and complications trial (DCCT) // Diabetes care. 1988. Vol. 11, № 9. P. 725–732.
4. Toobert D. J., Glasgow R. E. Assessing diabetes self-management: the summary of diabetes self-care activities questionnaire // Handbook of psychology and diabetes: A guide to psychological measurement in diabetes research and practice. 1994. Vol. 351. P. 75.
5. Omar K. et al. Validation of the Malay Version of Adolescent Coping Scale // International Medical Journal. 2011. Vol. 18, № 4.
6. Jalaludin M. Y. et al. Reliability and validity of the Malay translated version of diabetes quality of life for youth questionnaire // Malaysian family physician: the official journal of the Academy of Family Physicians of Malaysia. 2013. Vol. 8, № 1. P. 12.
7. Bujang M. A. et al. Validation of the summary of diabetes self-care activities (SDSCA) in Malay language for Malaysian adults // MJPHM. 2016. Vol. 16, № 3. P. 227–234.
8. Guillemin F., Bombardier C., Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines // Journal of clinical epidemiology. 1993. Vol. 46, № 12. P. 1417–1432.
9. De Vet H. C., Terwee C. B., Knol D. L., Bouter L. M. When to use agreement versus reliability measures // J Clin Epidemiol. 2006. Vol. 59, No. 10. P. 1033–1039.
12. Tavakol M., Dennick R. Making sense of Cronbach's alpha // International journal of medical education. 2011. Vol. 2. P. 53.
10. Tan S. M. K. et al. Validation of the Malay Version of Children's Depression Inventory // International Medical Journal. 2013. Vol. 20, № 2.
11. Bartko J. J. The intraclass correlation coefficient as a measure of reliability // Psychological reports. 1966. Vol. 19, № 1. P. 3–11.
12. Curtis P. S. et al. Extraction and critical appraisal of data // Handbook of Meta-analysis in Ecology and Evolution. 2013. P. 52–60.



13. Streiner D. L., Norman G. R., Cairney J. Health measurement scales: a practical guide to their development and use. Oxford University Press, USA, 2015.
14. Bujang M. A., Baharum N. Guidelines of the minimum sample size requirements for Kappa agreement test // *Epidemiology, Biostatistics and Public Health*. 2017. Vol. 14, № 2.
15. Zumbo B. D., Gadermann A. M., Zeisser C. Ordinal versions of coefficients alpha and theta for Likert rating scales // *Journal of modern applied statistical methods*. 2007. Vol. 6, № 1. P. 4.
16. Gadermann A. M., Guhn M., Zumbo B. D. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide // *Practical Assessment, Research, and Evaluation*. 2019. Vol. 17, № 1. P. 3.
17. McDonald R. P. Test theory: A unified treatment. Psychology press, 2013.
18. DeVellis R. F., Thorpe C. T. Scale development: Theory and applications. Sage publications, 2021.
19. Raykov T., Marcoulides G. A. Introduction to psychometric theory. Routledge, 2011.
20. Revelle W. Hierarchical cluster analysis and the internal structure of tests // *Multivariate Behavioral Research*. 1979. Vol. 14, № 1. P. 57–74.
21. Revelle W., Zinbarg R. E. Coefficients alpha, beta, omega, and the glb: Comments on Sijsma // *Psychometrika*. 2009. Vol. 74. P. 145–154.
22. Lance C. E., Butts M. M., Michels L. C. The sources of four commonly reported cutoff criteria: What did they really say? // *Organizational research methods*. 2006. Vol. 9, № 2. P. 202–220.
23. Nunnally J. C., Bernstein I. H. Psychometric Theory. New York, NY: McGraw-Hall. 1994.
24. Cronbach L. J. Coefficient alpha and the internal structure of tests // *Psychometrika*. 1951. Vol. 16, № 3. P. 297–334.
25. Lawshe C. H. et al. A quantitative approach to content validity // *Personnel psychology*. 1975. Vol. 28, № 4. P. 563–575.
26. Bujang M. A., Omar E. D., Baharum N. A. A review on sample size determination for Cronbach's alpha test: a simple guide for researchers // *The Malaysian journal of medical sciences: MJMS*. 2018. Vol. 25, № 6. P. 85.
27. Gravetter F. J., Forzano L. A. B. Research methods for the behavioral sciences. Cengage learning, 2018.
28. Davidson C. L. et al. Psychometric properties of the co-rumination questionnaire // *Personality and Individual Differences*. 2014. Vol. 70. P. 171–175.

29. Tiong X. T. et al. Validation of the Kessler's Psychological Distress Scale (K10 & K6) in a Malaysian Population // ASEAN Journal of Psychiatry. 2018. Vol. 19, No. 1.
30. Jacobson A. M., Groot M. D., Samson J. A. The evaluation of two measures of quality of life in patients with type I and type II diabetes // Diabetes care. 1994. Vol. 17, № 4. P. 267–274.
31. Yong A. G. et al. A beginner's guide to factor analysis: Focusing on exploratory factor analysis // Tutorials in quantitative methods for psychology. 2013. Vol. 9, № 2. P. 79–94.
32. Lovibond S. H. Manual for the depression anxiety stress scales // Sydney psychology foundation. 1995.
33. Daza P. et al. The depression anxiety stress scale-21: Spanish translation and validation with a Hispanic sample // Journal of psychopathology and behavioral assessment. 2002. Vol. 24. P. 195–205.
34. Musa R. et al. Translation, validation and psychometric properties of Bahasa Malaysia version of the Depression Anxiety and Stress Scales (DASS) // ASEAN Journal of Psychiatry. 2007. Vol. 8, № 2. P. 82–89.

## ПРИЛОЖЕНИЕ А

Коэффициент внутриклассовой корреляции – описательная статистика, представляющая собой отношение дисперсии наблюдений к дисперсии ошибок (1):

$$ICC = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\varepsilon}^2}, \quad (1)$$

где  $\sigma_{\alpha}^2$  – дисперсия наблюдений,  $\sigma_{\varepsilon}^2$  – дисперсия ошибок.

Если дисперсия ошибок равна или превышает интересующую дисперсию (например, дисперсию между испытуемыми), надежность метода низкая. Будет ли значение ICC достаточно хорошим, будет зависеть от используемой модификации ICC. Зависимость модификаций формулы ICC от допущений, принимаемых при планировании исследования, рассмотрена в таблице А.1.

Например, в случае, когда каждый из трех оценщиков дважды измерил всех 15 испытуемых, и для целей исследования важно не точное значение измерения, а лишь взаиморасположение 15 испытуемых между собой на основании произведенных измерений, при этом важно сохранить информацию о возможной разнице между измерителями, подходящей моделью будет ICC2 (С,1).

Функции для расчета коэффициента внутриклассовой корреляции доступны в составе статистических пакетов языков программирования R (пакеты *performance* 0.10.8, *irr* 0.84.1) и Python (библиотека *pengouin* 0.5.3), а также в составе программного обеспечения Stata и SPSS [13].

Таблица А.1 – Зависимость модификаций формулы ИСС от допущений, принимаемых при планировании исследования

Односторонний или двусторонний коэффициент: ИСС 1/2 (n, m)	Оценка согласованности или согласия: ИССх (С/А, m)	Единичный или средний коэффициент: ИССх (n, 1/k)
ИСС 1 (n, m): оценщики назначаются испытуемым случайным образом	ИССх (С, m): согласованность – совпадение относительного расположения измерений друг относительно друга между разными оценщиками	ИССх (n, 1): надежность оценивается для случая единственного измерения испытуемого оценщиком
ИСС 2 (n, m): все оценщики оценивают всех испытуемых	ИССх (А, m): согласие – абсолютное согласие измерений между оценщиками	ИССх (n, k): надежность оценивается для случая усреднения измерений испытуемого k-числом оценщиков
<p><i>Примечание. Согласованность (С) ИСС не может быть оценена в односторонней модели, поскольку нет способа разделить промежуточные и остаточные отклонения.</i></p>		

## ПРИЛОЖЕНИЕ Б

Каппа Коэна – коэффициент, предложенный Коэном в 1960 году для оценки согласованности между наблюдениями с учетом того, что часть этой согласованности может быть случайной. Используется только для категориальных данных. Для категориальных наблюдений номинального типа используется стандартная каппа Коэна, для порядковых – взвешенная каппа Коэна.

Формула для расчета стандартной каппы Коэна (2):

$$k = \frac{p_o - p_e}{1.0 - p_e}, \quad (2)$$

где  $p_o$  – наблюдаемое согласие между оценщиками,  $p_e$  – вероятность случайного согласия, использующая для вычисления вероятности наблюдаемые данные.

Формула для расчета взвешенной каппы Коэна (3):

$$k_w = 1.0 - \frac{\sum w_{ij} \times P_{oij}}{\sum w_{ij} \times P_{eij}}, \quad (3)$$

где  $w_{ij}$  – значения веса для каждой разницы между вариантами,  $P_{oij}$  и  $P_{eij}$  – наблюдаемые и ожидаемые согласия между оценщиками.

Функции для расчета коэффициента каппа Коэна доступны в составе статистических пакетов языков программирования R (пакеты *DescTools* 0.99.51, *psych* 2.3.6) и Python (библиотека *scikit-learn* 1.3.6), а также в составе программного обеспечения Stata и SPSS [13].

## ПРИЛОЖЕНИЕ В

Коэффициент альфа Кронбаха – статистика, оценивающая внутреннюю согласованность характеристик, описывающих один объект.

Формула для расчета альфы Кронбаха (4):

$$\alpha = \frac{k\bar{c}}{\bar{v} + (k-1)\bar{c}}, \quad (4)$$

где  $k$  – количество оцениваемых характеристик,  $\bar{c}$  – средняя ковариация между характеристиками,  $\bar{v}$  – средняя дисперсия характеристик.

Функции для расчета коэффициента альфа Кронбаха доступны в составе статистических пакетов языков программирования R (пакет *psych* 2.3.6) и Python (библиотека *pingouin* 0.5.3), а также в составе программного обеспечения Stata и SPSS [13].

*Серия «Лучшие практики лучевой и инструментальной диагностики»*

Выпуск 133

**Составители:**

*Васильев Юрий Александрович  
Владимирский Антон Вячеславович  
Мнацаканян Марина Генриковна  
Омелянская Ольга Васильевна  
Решетников Роман Владимирович  
Алымова Юлия Александровна  
Шумская Юлия Федоровна  
Ахмедзянова Дина Альфредовна*

## **МЕТОДИКА ВАЛИДАЦИИ СРЕДСТВ МЕДИЦИНСКОГО АНКЕТИРОВАНИЯ (ОПРОСНИКОВ)**

Методические рекомендации

Отдел координации научной деятельности ГБУЗ «НПКЦ ДиТ ДЗМ»  
Технический редактор В. П. Гамарина  
Компьютерная верстка Е. Д. Бугаенко

ГБУЗ «НПКЦ ДиТ ДЗМ»  
127051, г. Москва, ул. Петровка, д. 24, стр. 1