

The background features two stylized trees. The tree on the left is green with many small leaves, while the tree on the right is black with sparse red and yellow leaves, suggesting autumn. Numerous small leaves are scattered throughout the air, falling from the trees. The overall color palette is light green and beige.

КАЧЕСТВЕННАЯ КЛИНИЧЕСКАЯ ПРАКТИКА

www.ClinVest.ru

№2 2015



- Независимая научная организация в области изучения экономики программ здравоохранения
- Более 190 исследований в области оценки медицинских технологий
- Опубликовано около 200 научных работ в рецензируемых медицинских журналах
- Партнёры – 40 ведущих зарубежных и российских фармацевтических компаний

Основные направления научной деятельности

- разработка научно-методических основ фармакоэкономических, фармакоэпидемиологических исследований и оценки качества жизни
- развитие методологии проведения оценки медицинских технологий

ПОЛНЫЙ СПЕКТР УСЛУГ ПО ЭКОНОМИЧЕСКОЙ ОЦЕНКЕ

- фармакоэкономические исследования
- фармакоэпидемиологические исследования
- наблюдательные неинтервенционные исследования
- непрямые сравнительные исследования
- оценка технологий здравоохранения
- оценка качества жизни, связанного со здоровьем
- систематический литературный обзор и мета-анализ
- разработка математических моделей и локальная адаптация
- анализ больших баз данных
- создание мобильных приложений и интерактивных онлайн-презентаций
- формирование доказательной базы
- экспертиза и разработка клинико-экономического досье
- образовательные услуги
- информационно-консультационные услуги





Издательство

ОКИ

**ГЛАВНЫЙ РЕДАКТОР
Белоусов Ю.Б.**

**Зам. главного редактора
Леонова М.В.**

Редакционная коллегия

Аверков О.В.	Лазебник Л.Б.
Акопян А.С.	Майорова О.А.
Арутюнов Г.П.	Малаев М.Г.
Ашихмин Я.И.	Маликов А.Я.
Барт Б.Я.	Морозова Т.Е.
Батищева Г.А.	Мохов О.И.
Батурин В.А.	Мухин К.Ю.
Бондарева И.Б.	Незванов Н.Г.
Быков А.В.	Никанорова М.Ю.
Вёрткин А.Л.	Никитин И.Г.
Власов П.Н.	Огородова Л.М.
Вольская Е.А.	Омельяновский В.В.
Гельфанд Б.Р.	Прохорович Е.А.
Грацианский Н.А.	Самсыгина Г.А.
Гуревич К.Г.	Семернин Е.Н.
Гучев И.А.	Сидоренко С.В.
Егоров Е.А.	Соколов А.В.
Звартау Э.Э.	Сычёв Д.А.
Зырянов С.К.	Трофимов В.И.
Карпов О.И.	Фролов М.Ю.
Колбин А.С.	Хохлов А.Л.
Кобалава Ж.Д.	Шляхто Е.В.
Котовская Ю.В.	Щербаков П.Л.
Кошкин В.М.	Явелов И.С.

Издательская группа:

Научный редактор: Белоусов Д.Ю.
Дизайн, верстка: Design2pro.ru
Подписано в печать: 23.10.2015.
Типография: ООО «Тверской Печатный Двор»
г. Тверь, с. Никольское, д. 26
Тираж: 3700 экз.
Свободная цена.

Учредитель: Общество клинических исследователей

Журнал зарегистрирован Комитетом РФ
по печати 28.05.2001 г.
Свидетельство о регистрации средства
массовой информации ПИ №779142.
Оформить подписку можно
через «Агентство «Книга-Сервис» или каталог
«Пресса России» — **подписной индекс 45071**

Авторские материалы не обязательно
отражают точку зрения редакции.
Редакция не несет ответственности
за достоверность информации,
содержащейся в рекламных материалах.

e-mail: clinvest@mail.ru
www.zdat-oki.ru; www.clinvest.ru
Тел.: +7 (910) 449-22-73

Сайты

PharmacoKinetics.ru
ClinVest.ru
Clinical-Pharmacy.ru
PharmacoGenetics-PharmacoGenomics.ru
Antibiotics-Chemotherapy.ru

Журналы

Фармакокинетика и Фармакодинамика
Качественная клиническая практика
Клиническая фармация
Фармакогенетика и Фармакогеномика
Антибиотики и Химиотерапия

WEB-порталы

HealthEconomics.ru
Market-Access-Solutions.ru
zdat-oki.ru
Центр фармакоэкономических исследований
Market Access Solutions
Издательство ОКИ

СОДЕРЖАНИЕ

КЛИНИЧЕСКИЕ ИССЛЕДОВАНИЯ

Валидация электронных систем сбора данных «Оценок Результатов Пациентами» — Рекомендации для клинических исследователей: отчёт Целевой Группы ISPOR по надлежащей исследовательской практике валидации ePRO систем
Arthur Zbrozek, Joy Hebert, Gregory Gogates, Rod Thorell, Christopher Dell, Elizabeth Molsen, Gretchen Craig, Kenneth Grice, Scottie Kern, Sheldon Hines 3

ОЦЕНКА МЕДИЦИНСКИХ ТЕХНОЛОГИЙ

Принципы надлежащей практики для применения аналитического моделирования в оценке медицинских технологий: доклад Рабочей группы ISPOR по надлежащей практике исследований — исследования по моделированию
Milton C. Weinstein, Bernie O'Brien, John Hornberger, Joseph Jackson, Magnus Johannesson, Chris McCabe, Bryan R. Luce 19

Многонациональные исследования — Рекомендации по требованиям к переводам, подходам к использованию одного языка в разных странах, а также к подходам в поддержке при сборе данных: отчет Целевой группы ISPOR по надлежащей исследовательской практике в переводах и лингвистической валидации оценок результатов пациентами
Diane Wild, Sonya Eremenco, Isabelle Mear, Mona Martin, Caroline Houchin, Mary Gawlicki, Asha Hareendran, Ingela Wiklund, Lee Yee Chong, Robyn von Maltzahn, Lawrence Cohen, Elizabeth Molsen..... 29

Инструменты для сбора данных сообщений об исходах от медицинского вмешательства от пациентов (PRO) применяемые в педиатрии и предназначенные для инструкции для медицинского продукта: Отчёт Исследовательской группы ISPOR по разработке Надлежащих Исследовательских Практик о сообщаемых пациентами исходах
Louis S. Matza, Donald L. Patrick, Anne W. Riley, John J. Alexander, Luis Rajmil, Andreas M. Pleil, Monika Bullinger 44

КАЧЕСТВО ЖИЗНИ

QALY: основы
Milton C. Weinstein, George Torrance, Alistair McGuire 70

Сохранить и улучшить QALY
Joseph Lipscomb, Michael Drummond, Dennis Fryback, Marthe Gold, Dennis Revicki 79

Применение QALY в клинической практике и при принятии административных решений: проблемы и перспективы
Paul Kind, Jennifer Elston Lafata, Karl Matuszewski, Dennis Raisch..... 92

Слово к Консенсусу по QALY
Michael Drummond, Diana Brixner, Marthe Gold, Paul Kind, Alistair McGuire, Erik Nord (Consensus Development Group) 98

АНАЛИЗ ВЛИЯНИЯ НА БЮДЖЕТ

Анализ влияния на бюджет: принципы надлежащей практики. Доклад рабочей группы ISPOR по надлежащей практике анализа влияния на бюджет II, 2012 год
Sean D. Sullivan, Josephine A. Mauskopf, Federico Augustovski, J. Jaime Caro, Karen M. Lee, Mark Minchin, Ewa Orlewska, Pete Penna, Jose-Manuel Rodriguez Barrios, Wen-Yi Shau 104

**EDITOR-IN-CHIEF
Belousov U.B.**

**Deputy Editor-In-Chief
Leonova M.V.**

Editorial Board

Averkov O.V.	Lazebnik L.B.
Akopyan A.S.	Mayorov O.A.
Arutyunov G.P.	Malaev M.G.
Ashikhmin Y.I.	Malikov A.J.
Bart B.J.	Morozova T.E.
Batishcheva G.A.	Mokhov O.I.
Baturin V.A.	Mukhin K.Y.
Bondareva I.B.	Neznanov N.G.
Bykov A.V.	Nikanorova M.Y.
Vertkin A.L.	Nikitin I.G.
Vlasov P.N.	Ogorodova L.M.
Volskay E.A.	Omelyanovskii V.V.
Gelfand B.R.	Prokhorovich E.A.
Gratsianskii N.A.	Samsygina G.A.
Gurevich K.G.	Semernin E.N.
Guchev I.A.	Sidorenko S.V.
Egorov E.A.	Sokolov A.V.
Zvartau E.E.	Sychev D.A.
Ziryaynov S.K.	Trofimov V.I.
Karpov O.I.	Frolov M.U.
Kolbin A.S.	Khokhlov A.L.
Kobalava J.D.	Shlyakhto E.V.
Kotovskaya Y.V.	Shcherbakov P.L.
Koshkin V.M.	Yavelov I.S.

Publishing Group:

Scientific editor: Belousov D.U.
Design by Design2pro.ru
Signed in print: 23.10.2015
Printing house: Ltd «Tver Pechatniy Dom»,
Tver, s. Nikolskoe, 26
Circulation: 3700 copies. Free price.

Founder: Russian society of clinical investigators

The journal is registered by the Committee of the Russian Federation of Press 28.05.2001
Certificate of media registration ПИ №779142.

You can subscribe via the «Agency «Book-Service»
or the catalog «Press of Russia» — **Index 45071**

Copyright material does not necessarily
reflect the views of the publisher.
We take no responsibility for the information contained
in promotional materials.

LLC «Publisher OCI»

e-mail: clininvest@mail.ru
www.izdat-oki.ru; www.clininvest.ru
Tel.: +7 (910) 449-22-73

Sites

PharmacoKinetica.ru
ClinVest.ru
Clinical-Pharmacy.ru
PharmacoGenetics-PharmacoGenomics.ru
Antibiotics-Chemotherapy.ru

Journals

Pharmacokinetics and Pharmacodynamics
Good Clinical Practice
Clinical Pharmacy
Pharmacogenetics and Pharmacogenomics
Antibiotics and Chemotherapy

WEB-portals

HealthEconomics.ru
Market-Access-Solutions.ru
Izdat-Oki.ru
Center of Pharmacoeconomics Research
Market Access Solutions
Publisher OCI

CONTENTS

CLINICAL STUDIES

Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data — Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems Validation Good Research Practices Task Force
Arthur Zbrozek, Joy Hebert, Gregory Gogates, Rod Thorell, Christopher Dell, Elizabeth Molsen, Gretchen Craig, Kenneth Grice, Scottie Kern, Sheldon Hines..... 3

HEALTH TECHNOLOGY ASSESSMENT

Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices — Modeling Studies
Milton C. Weinstein, Bernie O'Brien, John Hornberger, Joseph Jackson, Magnus Johannesson, Chris McCabe, Bryan R. Luce..... 19

Multinational Trials — Recommendations on the Translations Required, Approaches to Using the Same Language in Different Countries, and the Approaches to Support Pooling the Data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force Report
Milton C. Weinstein, Bernie O'Brien, John Hornberger, Joseph Jackson, Magnus Johannesson, Chris McCabe, Bryan R. Luce..... 29

Pediatric Patient-Reported Outcome Instruments for Research to Support Medical Product Labeling: Report of the ISPOR Good Research Practices for the Assessment of Patient-Reported Outcomes in Children and Adolescents Task Force
Louis S. Matza, Donald L. Patrick, Anne W. Riley, John J. Alexander, Luis Rajmil, Andreas M. Pleil, Monika Bullinger..... 44

QALY

QALYs: The Basics
Milton C. Weinstein, George Torrance, Alistair McGuire..... 70

Retaining, and Enhancing, the QALY
Joseph Lipscomb, Michael Drummond, Dennis Fryback, Marthe Gold, Dennis Revicki..... 79

The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects
Paul Kind, Jennifer Elston Lafata, Karl Matuszewski, Dennis Raisch..... 92

Toward a Consensus on the QALY
Michael Drummond, Diana Brixner, Marthe Gold, Paul Kind, Alistair McGuire, Erik Nord (Consensus Development Group)..... 98

BUDGET IMPACT ANALYSIS

Budget impact analysis: principles of good practice. Report of the ISPOR Working Group on Good Practices for Budget Impact Analysis II, 2012
Sean D. Sullivan, Josephine A. Mauskopf, Federico Augustovski, J. Jaime Caro, Karen M. Lee, Mark Minchin, Ewa Orlewska, Pete Penna, Jose-Manuel Rodriguez Barrios, Wen-Yi Shau..... 104

Валидация электронных систем сбора данных «Оценок Результатов Пациентами» — Рекомендации для клинических исследователей: отчёт Целевой Группы ISPOR по надлежащей исследовательской практике валидации ePRO систем

Arthur Zbrozek¹, Joy Hebert², Gregory Gogates³, Rod Thorell⁴, Christopher Dell⁵, Elizabeth Molsen⁶, Gretchen Craig⁷, Kenneth Grice⁸, Scottie Kern⁹, Sheldon Hines¹⁰

¹ — Global Health Economics, CSL Behring, Biotherapies for Life, King of Prussia, PA, USA

² — PHT Corp, Boston, MA, USA

³ — CRF Health, Plymouth Meeting, PA, USA

⁴ — PHT Corp, Charlestown, MA, USA

⁵ — ERT, Philadelphia, PA, USA

⁶ — ISPOR, Lawrenceville, NJ, USA

⁷ — ERT, Inc., Pittsburgh, PA, USA

⁸ — Global Electronic Data Capture, Bayer HealthCare Pharmaceuticals, Inc., Montville, NJ, USA

⁹ — Sacaja Consulting Ltd., Wiltshire, England, UK

¹⁰ — ICON Late Phase & Outcomes Research, Durham, NC, USA

Переводчики:

Павлыш Андрей Владиславович — к.м.н., докторант Первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Топоров Александр Андреевич — студент Первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Рецензенты:

Вербицкая Елена Владимировна — к.ф.-м.н., ассистент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Колбин Алексей Сергеевич — д.м.н., профессор, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Резюме. Литература об исследовании исходов лечения изобилует примерами высококачественных, надёжных данных оценок результатов пациентами (Patient-Reported Outcomes, PRO), введённых исключительно электронными методами, ePRO, которые сравниваются с данными, вводимыми с «бумажных носителей». Менеджеры клинических исследований всё чаще используют сбор данных с помощью ePRO для получения результатов, основанных на PRO. Проверка регуляторных документов позволяет определить правила, которым необходимо следовать, чтобы собирать данные с помощью ePRO в соответствии с требованиями ведения медицинской документации. Критический компонент соблюдения нормативных требований — это свидетельство о валидации (evidence of the validation) системы сбора электронных данных. Валидация (validation) электронных систем — это процесс в отличие от сконцентрированной деятельности (focused activity), которая совершается в определённый момент времени. Валидация программного продукта для сбора данных предусматривает описание и выполнение в условиях его целевой среды следующих восьми шагов: описание требований к системе (requirements definition), дизайн (design), кодирование (coding), тестирование (testing), прослеживание процесса (tracing), тестирование на приемлемость для пользователя (user acceptance testing), установка (installation) и конфигурирование системы (configuration), а также списание (decommissioning). Указанные элементы согласуются с последними руководствами регуляторных органов по валидации систем. Этот отчёт был написан, чтобы объяснить, каким образом процесс валидации работает для спонсоров, клинических исследователей и прочих пользователей устройств по сбору электронных данных, ответственных за проверку качества данных, вводимых в реляционные базы данных посредством подобных устройств. Это руководство устанавливает 2 документальных требования, необходимые системному провайдеру системы сбора данных для демонстрации валидности системы. Это практический источник информации для исследовательских групп, позволяющий убедиться, что поставщики ePRO используют те способы валидации системы и реализации процессов, которые гарантируют, что системы и сервисы работают надёжно на практике, производят точные, полные и целостные данные (файлы данных), поддерживают административное управление и соответствуют текущим регуляторным требованиям. Более того, этот короткий отчёт улучшит понимание требований пользователя к технической валидации, что ведёт к более информированным и сбалансированным рекомендациям или решениям по вопросам методов сбора электронных данных.

Ключевые слова: сбор электронных данных, ePRO, PRO, валидация системы

Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data — Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems Validation Good Research Practices Task Force

Abstract. Outcomes research literature has many examples of high-quality, reliable patient-reported outcome (PRO) data entered directly by electronic means, ePRO, compared to data entered from original results on paper. Clinical trial managers are increasingly using ePRO data collection for PRO-based endpoints. Regulatory review dictates the rules to follow with ePRO data collection for medical label claims. A critical component for regulatory compliance is evidence of the validation of these electronic data collection systems. Validation of electronic systems is a process versus a focused activity that finishes at a single point in time. Eight steps need to be described and undertaken to qualify the validation of the data collection software in its target environment: requirements definition, design, coding, testing, tracing, user acceptance testing, installation and configuration, and decommissioning. These elements are consistent with recent regulatory guidance for systems validation. This report was written to explain how the validation process works for sponsors, trial teams, and other users of electronic data collection devices responsible for verifying the quality of the data entered into relational databases from such devices. It is a guide on the requirements and documentation needed from a data collection systems provider to demonstrate systems validation. It is a practical source of information for study teams to ensure that ePRO providers are using system validation and implementation processes that will ensure the systems and services: operate reliably when in practical use; produce accurate and complete data and data files; support management control and comply with any existing regulations. Furthermore, this short report will increase user understanding of the requirements for a technology review leading to more informed and balanced recommendations or decisions on electronic data collection methods.

Keywords: electronic data collection, ePRO, PRO, systems validation

Опубликовано: Arthur Zbrozek, Joy Hebert, Gregory Gogates, Rod Thorell, Christopher Dell, Elizabeth Molsen, Gretchen Craig, Kenneth Grice, Scottie Kern, Sheldon Hines. Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data — Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems Validation Good Research Practices Task Force. // Value in Health 16 (2013) 480-489.

История создания Целевой Группы

Целевая Группа ISPOR по надлежущей исследовательской практике валидации ePRO-систем была сформирована из созданной ранее рабочей группы по данному вопросу и утверждена советом директоров ISPOR в марте 2011 года. Руководящая часть

Целевой Группы состояла из экспертов в области электронных систем сбора данных с опытом работы в разработке дизайна, контроле качества и регуляторных вопросах, а также имеющих опыт работы в клинических исследованиях. Группа встречалась дважды в месяц для разработки общей схемы достижения конечной цели: информирования пользовате-

лей о требованиях к уровню качества и содержанию валидированных систем сбора данных. Авторы работали группами и поодиночке над разными частями отчёта, которые позже были рассмотрены Целевой Группой в полном составе для комментариев и замечаний.

Как только был подготовлен первый вариант отчёта, его отправили на ознакомление 400-м членам наблюдательной группы ISPOR PRO и представителю FDA (Управления по контролю качества пищевых продуктов и лекарственных средств), знакомому с темой. Кроме того, к этому моменту работа была представлена для комментариев в виде презентации на 16-м конгрессе ISPOR в Балтиморе, США. Члены ISPOR внесли свой вклад в этот отчёт, добавив письменные комментарии во время ознакомления и устные комментарии во время презентации. Авторы ещё несколько раз вычитывали отчёт и отправили финальную версию членам наблюдательной группы ISPOR PRO, одновременно с этим пригласив к ознакомлению ISPOR в полном составе.

Все комментарии, большинство из которых оказались существенными и конструктивными, были рассмотрены основателями Целевой Группы и сочтены уместными. Дальнейшие дополнения были внесены благодаря полученным отзывам, и как только все авторы между собой пришли к соглашению, окончательный вариант отчёта был опубликован в журнале «Value in Health».

Письменные комментарии и список рецензентов опубликованы на сайте ISPOR на странице Целевой Группы: <http://www.ispor.org/sigs/ePROsystemvalidationsg.asp>. Отчёт Целевой Группы и интернет-страницу можно также найти на домашней странице ISPOR (www.ispor.org), выбрав лиловое меню «Research Tools» (Инструменты исследования), а в нём раздел «Good Practices for Outcomes Research» (Надлежащая практика изучения исходов).

Введение

Опыт пациентов в оценке эффективности и безопасности медицинских продуктов, в особенности лекарственных средств и медицинских изделий, приобретает всё большее значение. Он дополняет применение клинических оценок, объективную статистику, к примеру, долю выживаемости, и прочие традиционные индикаторы клинической действенности и безопасности. Клинические исследователи постоянно включают «оценки исходов по отчётам пациентов» (**«оценки результатов пациентами» (Patient-Reported Outcomes, PRO)**) в программу клинических исследований, поскольку это помогает измерить эффект медицинского продукта на такие факторы как тяжесть симптома и физическая или психическая функция. Данные PRO могут быть первичной или вторичной конечной точкой в

определении действенности лечения. В некоторых случаях, таких как оценка утомляемости или боли, PRO может оказаться единственной реальной конечной точкой, поскольку в этих случаях не существует маркеров болезни или активности лечения, которые смог бы измерить исследователь, наблюдатель или лаборатория [1].

Согласно определению FDA, PRO — это «любой отчёт о состоянии здоровья пациента, который исходит напрямую от пациента без интерпретации ответа пациента врачом или кем-либо ещё» [2]. Измерения можно проводить как в абсолютных (например, тяжесть признака, симптома или состояния здоровья), так и в относительных значениях — в рамках изменений со времени последнего измерения [2]. В статье ЕМА (Европейское Медицинское Агентство), отражающей регуляторные требования к использованию Шкал Оценки Качества Жизни, PRO определяется как «любой исход, напрямую оценивающийся пациентом и основывающийся на восприятии пациентом болезни и её лечения» [3]. Проще говоря, PRO — это последствия болезни и её лечения, о которых докладывает пациент [1].

В клинических исследованиях для органов контроля следующим по важности после вопроса о безопасности пациента является вопрос о качестве данных и их достоверности [4]. С точки зрения спонсора клинического исследования, целостность и качество данных имеют решающее значение в вопросе надёжности исследования, равно как и соблюдение требований FDA, ЕМА и прочих надзорных органов. Готовность FDA принять данные клинических исследований в целях принятия регуляторных решений зависит от их способности подтвердить качество и целостность данных во время инспекций и аудитов FDA [5].

Менеджеры клинических исследований всё чаще используют **ePRO (электронный сбор данных PRO напрямую от пациентов)** для формирования конечных точек на основе PRO. ePRO ведёт к улучшению качества и полноты данных, снижению субъективного и административного влияния, так же обработке алгоритмов перехода внутри данных (skip patterns) [1, 6]. Электронный сбор данных приносит более надёжную и точную информацию, позволяет проводить более глубокое тестирование объектов исследований и позволяет увидеть более правильную картину опыта пациента [6]. Регуляторные документы определяют правила, по которым должен проходить электронный сбор данных. Вне зависимости от того, использует ли специалист, проводящий исследование, электронный или бумажный опросник для сбора данных, основные принципы, касающиеся точности данных (например, возможность оперативного контроля и отслеживаемость изменений) необходимы к соблюдению как в электронных, так и в бумажных версиях.

Для демонстрации того, что субъекты интерпретируют и отвечают на вопросы/элементы PRO одинаково вне зависимости от способа сбора данных, требуются доказательства [1, 5]. Процедура изменения способа сбора данных, а также оценка эквивалентности измерений обоими способами уже были описаны в предыдущем отчёте Целевой Группы ISPOR по PRO: «Рекомендации по мерам предоставления доказательств, необходимых для подтверждения эквивалентности измерений оценок результатов пациентами в электронном и бумажном виде: отчёт Целевой Группы ISPOR по надлежащей исследовательской практике в ePRO» [1].

Вне зависимости от типа администрирования (самостоятельно или с помощью интервьюера) или метода электронного сбора данных (приспособлений, использующихся для фиксирования данных: система интерактивного голосового управления (interactive voice response systems), ввод данных через интернет-портал (Web-based data entry) или устройство ePRO (ePRO devices)), валидация системы должна соответствовать стандартам FDA и EMA. Это происходит во время валидации процесса разработки, поддержки и работы устройства и компьютерной системы [5–7]. Проще говоря, должны существовать доказательства, что процесс идёт так, как должен. Например, если на экран с помощью устройства ввода данных или сенсорно вводится число «5», ответ субъекта должен точно соответствовать значению из базы данных — правильно фиксируясь как «5» в самой базе.

В случае с ePRO это осложняется тем фактом, что существующие правила и руководства изначально разрабатывались для бумажных опросников и дневников. В настоящий момент единого метода разработки и внедрения, предписанного к использованию регуляторами этой сферы, нет. В связи с отсутствием специальных правил или руководств от этих органов о том, как именно должна проводиться валидация систем ePRO, было решено опираться на похожие стандарты из других руководств по схожим темам, например, таким как валидация систем, используемых для разработки медицинских устройств. (Дополнительные материалы о правилах в отношении клинических исследований, разработки и валидации систем ePRO можно найти по ссылке <http://dx.doi.org/10.1016/j.jval.2013.04.002>.)

Существует два способа поставки **программно-го обеспечения (ПО)** для ePRO, каждый из которых нуждается в собственном методе валидации. Первый — это традиционный метод заказного ПО — для каждого клинического исследования разрабатывают отдельное ПО. Допустимо использование частей кода существующего ПО. Вся система проходит через набор жёстких валидационных тестов, прежде чем ее устанавливают для использования в исследовании [8]. Второй вариант — это платформа,

созданная подрядчиком, которую адаптируют и перенастраивают для каждого исследования. Платформа такого рода проходит весь набор строгих валидационных тестов во время изначальной разработки. Этот метод позволяет валидировать только процедуры адаптации платформы к конкретному исследованию. Оба метода поставки имеют свои преимущества. Один позволяет сохранять полную гибкость, но за большую стоимость, и процесс разработки требует больше времени; в то время как другой позволяет ускорить разработку и снизить затраты, за счёт ограничения гибкости системы.

Основная цель этого отчёта — помочь понять техническую природу систем ePRO и процесса проверки систем ePRO, в связи с тем, что техническая сторона валидации систем ePRO и регуляторные требования к валидации могут быть не вполне ясны спонсорам проекта и членам исследовательской группы. Процесс валидации ePRO определяет ответственность, которую делят между собой системный провайдер ePRO и спонсор, который использует ePRO. Важно понимать природу этой ответственности, и каким образом она должна быть разделена.

Вторая цель этого отчёта — создать рекомендации для спонсоров и менеджеров проекта исследования по валидации системы, обратив особое внимание на обязанности, которые берёт на себя каждый участник. Этот отчёт посвящён технической природе систем электронного сбора данных ePRO и процессу валидации. Он поможет спонсору понять требования, предъявляемые к технологической проверке и основаниям для сравнения поставщиков разных систем ePRO и, соответственно, к их предложениям об услугах и давать более сбалансированные рекомендации для принятия решений по вопросам систем электронного сбора данных.

Кроме того, данный отчёт поможет в понимании усилий, требуемых от спонсора, чтобы дополнить процесс валидации, которые предлагает системный провайдер. Когда системный провайдер ePRO просто предлагает устройство и систему данных, но не предлагает своих услуг по её валидации, груз выполнения этих обязательств ложится на спонсора. Заметим, что в настоящем отчёте термин «спонсор» относится к группе клинического исследования, работающей с поставщиком ePRO. Эти рекомендации пригодятся в равной мере контрактно-исследовательской организации (CRO) и любой другой организации, сотрудничающей с системным провайдером ePRO, чтобы создать систему ePRO, которую можно будет использовать в предрегистрационных исследованиях медицинских продуктов. Общие принципы, изложенные в этом отчёте, могут быть применены и к другим исследованиям, в которых субъекты используют электронные средства для ввода данных в виде ответов на вопросы.

И наконец, в отчёт включены подходящие регуляторные правила, так как в данном случае требуется

строгое соответствие стандартам в этих процессах и во время проведения самого исследования. Приложение 1 в Дополнительных материалах, которые можно найти по ссылке <http://dx.doi.org/10.1016/j.jval.2013.04.002>, включает в себя международные стандарты проведения клинических исследований, а также основные правила и инструкции FDA и ЕМА относительно разработки систем ePRO и их валидации для клинических исследований.

Основные принципы и определения валидации

В Руководстве FDA по общим принципам процесса валидации от 1987 г. содержится следующее определение *валидации*:

- установление документального свидетельства, подтверждающего высокий уровень уверенности, что данный конкретный процесс на постоянной основе производит продукт, который удовлетворяет всем предварительно заданным спецификациям и атрибутам качества [9].

Руководство 2011 года «Отраслевой процесс валидации: общие принципы и практика», подтверждает эти основные принципы и вносит в них изменения, соответствующие XXI-му веку. В настоящее время определение *валидации* выглядит так:

- сбор и проверка данных, с момента разработки концепции до выхода коммерческого продукта, который является научным свидетельством, что процесс способен на постоянной основе производить качественный продукт. Процесс валидации состоит из серии мероприятий, предпринимающихся на протяжении всего времени существования продукта и процесса [10].

Процесс валидации системы охватывает полный жизненный цикл разработки системы, от возникновения идеи до разработки, тестирования и релиза, с учётом последнего этапа — вывода из эксплуатации (списания). Согласно Руководству по надлежащей производственной практике (Good Manufacturing Practice, GMP) в ЕС, компьютеризованное руководство к системе, документация о валидации и отчёты должны описывать каждый шаг жизненного цикла системы. Производители ПО (в данном случае поставщики ePRO), должны иметь возможность подтвердить свои стандарты, протоколы и критерии приемлемости процедур и записей на основе оценки рисков [11]. Руководство FDA говорит о том же, рекомендуя объединить управление жизненным циклом ПО и управление рисками с проверкой и верификацией действий, предпринятых на протяжении всего жизненного цикла ПО [12].

Важно заметить, что валидацию не прекращают, когда ПО готово к выпуску или находится «на полке». Валидация — это не последний этап действий в

какой-то конкретный момент, а статус соответствия, поддерживаемый системой управления качеством на протяжении всего цикла жизни ПО [7, 12]. Чтобы поддерживать статус валидации, необходимо с вниманием относиться к документированию любых изменений и повторному исполнению любых необходимых валидационных процедур, пока система находится в действии — вплоть до момента вывода её из эксплуатации.

Валидация последовательно начинается с составления документации с описанием требований к системе и включает следующие шаги: дизайн (design), кодирование (coding), тестирование (testing), отслеживание исправления данных (tracing), тестирование на приемлемость для пользователя (user acceptance testing, UAT), установка (installation) и конфигурация (configuration), выведение из эксплуатации (decommissioning). Любые разработки ePRO должны соответствовать стандартному списку проектных документов, представленных в описанной ниже очередности.

Процесс валидации: основные составляющие. Жизненный цикл развития системы — это последовательность действий, задач, обязательств и отчётных материалов, необходимых для разработки высококачественной валидированной системы электронного сбора данных. Методология жизненного цикла развития системы, используемая поставщиком ePRO, должна быть тщательно изучена, чтобы определить, включает ли она в себя как минимум следующие критически важные элементы:

1. Требования системы (системные требования).
2. Дизайн системы.
3. Кодирование/адаптация/разработка ПО.
4. Трассировка (прослеживание).
5. Тестирование на приемлемость для пользователя.
6. Управление установкой/конфигурацией.
7. План вывода из эксплуатации.

Требования системы

Что это такое? Цель создания документации требований системы — описать все аспекты системы, вне зависимости от использованной технологии. Получившаяся документация покрывает нужды Протокола исследования, целевой группы пациентов и персонала медицинского учреждения. Её нужно будет формально утвердить до начала разработки системы [13].

Документация требований системы — это краткое содержание (план) того, что система будет выполнять. Она позволяет:

- 1) группе клинического исследования требовать внесения изменений в тот момент, когда ePRO существует только на бумаге — т.е., когда внести изменения можно быстро, малорискованно и недорого;

- 2) команде системного провайдера получить ясное и полное описание системы, которую потребуется спроектировать, разработать и произвести. Например, программистам, авторам стратегии тестирования и разработчикам нужна детальная картина каждого аспекта системы, чтобы выполнить свои обязательства по валидации системы и убедиться, что их собственные отдельные компоненты совпадают с компонентами других членов команды должным образом, что позволит разработать для исследования высококачественную систему;
- 3) как группой исследования, так и поставщиком ПО будут заданы одинаковые ожидания к системе, понятные в равной мере всем участникам. Это особенно важно при тестировании системы на приемлемость для пользователя (UAT) и использовании системы в исследовании (подробнее будет изложено ниже).

Прочие названия для этой документации.

Также этот раздел, помимо прочего, может называться «требованиями пользователя», «функциональными требованиями», «спецификацией пользователя». Документация может представлять собой один или несколько документов. Вне зависимости от её состава, для обеих сторон важно, чтобы в документации было предельно ясное и подробное описание всех аспектов системы [13, 14].

Вовлечённость группы клинического исследования

Участие этой группы необходимо на стадии составления требований и подбора определений, независимо от того, будет ли система ePRO разрабатываться внутренне или её будет разрабатывать внешний поставщик.

Те группы клинических исследований, у которых нет опыта написания документов по системным требованиям, для того, чтобы его написать, обычно, опираются на опыт системного провайдера ePRO, консультанта или технических экспертов в собственной организации. Группа клинического исследования предоставляет указанным экспертам Протокол клинического исследования. В большинстве случаев, группа встречается с авторами документа о системных требованиях, чтобы определить, как система должна работать, чтобы соответствовать требованиям Протокола, а также требованиям пользователей в клинических сайтах, целевой аудитории пациентов и заинтересованные стороны от организации-спонсора. Заинтересованные стороны от организации-спонсора включают клинические процедуры, менеджмент данных, биостатистику, исследование исходов или анализ стоимости лечения и соответствие нормативным требованиям [14].

С группой клинического исследования обсуждаются вопросы, которые включают в себя последова-

тельность экранных изображений (индивидуальный дизайн страниц для более точного отображения элементов PRO), изменение полей ввода данных, требования по переносу данных, формат файла данных, требования безопасности и соответствие Правилу 21 CFR часть 11 FDA об электронных записях [15], электронные подписи и множество других деталей. Спонсор должен перевести регуляторные требования на язык системной функциональности от доступа к данным до составления отчётов и сроке их хранения для аудита и проверок [6].

Когда предварительная версия документа о системных требованиях готова, необходимо, чтобы группа клинических исследований провела его вдумчивое рассмотрение и подготовила обратную связь, чтобы запросить любые необходимые изменения. Эти изменения по требованию спонсора должны быть внесены в соответствии с оговорёнными сроками, чтобы не менять срок поставки продукта. Если группа клинических исследований и системный провайдер ePRO удовлетворены документом о системных требованиях, считая, что он полностью и точно включает в себя все детали информации о системе, основные члены исследовательской группы подписывают этот документ, подтверждая таким образом своё согласие и удовлетворение. Основные члены команды системного провайдера также подписывают этот документ в знак подтверждения их согласия и готовности предоставить систему в описанном виде [13, 14].

Почему это важно? Документация о системных требованиях — это основа всех последующих процессов валидации системы и получаемых результатов. Если её разделы описаны неразборчиво, двояко или вовсе отсутствуют, их можно будет интерпретировать по-разному, что приведёт к недопониманию между группой клинического исследования и группой системного провайдера ePRO. Эта разница обнаружится только позже, на стадии тестирования на приемлемость для пользователя, что может закончиться задержкой при наборе пациентов, создаёт риск для обеспечения качества и стабильности ПО, и вызовет проблемы с функциональностью и недостатками работы ПО во время проведения исследования. Таким образом, крайне важно, чтобы этот документ был точным, ясным, правильным и полным [13, 14, 16].

Минимальное содержание

Документация с системными требованиями должна включать в себя всё необходимое для обеспечения чёткого, детального описания системы. Как минимум, они должны показывать и/или точно описывать все аспекты системы, влияющие на пользователей, включая выбранный инструментарий, снимки экрана (карманный персональный компьютер (КПК), планшет или другое сетевое устройство), голосовые подсказки (система интерактивного го-

лосового управления), поля ввода данных, проверку редактирования, сообщения об ошибках, логику навигации, поведение всех кнопок (КПК, планшета или другого сетевого устройства), алгоритмы, сигнализацию, тайм-ауты, функции безопасности, контрольный журнал, электронные подписи, предупреждения (сообщения о необходимости обратить внимание) субъектов, сотрудников сайта и дизайн отчёта. Текстовые описания этих элементов должны идти в сопровождении графики, если это необходимо для полной ясности.

Независимо от технологии, используемой в исследовании, идеальная документация с системными требованиями должна включать в себя следующие разделы:

- 1) цели и задачи;
- 2) определения;
- 3) справочные документы;
- 4) допущения;
- 5) системы и технологические потоки, в том числе блок-схемы, которые показывают последовательность экранов и связанную с ними логику всей системы;
- 6) функциональные требования. Последний раздел — «сердце» документа — может быть длиннее прочих, зачастую за него отвечает группа пользователей (например, Системные требования для сотрудников клиники), но может быть составлен в любой форме, которая делает его лёгким для понимания.

Краткое описание функциональных требований в документации о системных требованиях — пример того содержимого, которое стоит ожидать на первом экране в разделе, описывающем последовательность экранов для пациентов. За каждым элементом в структуре следует чёткое и подробное описание.

Функциональные требования

Раздел 1 — последовательность экранов для пациента

1. Первый экран — логин пациента
 - a. Снимок экрана
 - b. Данные на баннере экрана
 - 1) Дата
 - 2) Время
 - c. Поля ввода данных
 - PIN-код пациента
 - Проверка на 4-х-символьное цифровое сообщение об ошибке
 - Проверка, совпадает ли PIN-код и сообщение об ошибке
 - d. Кнопки навигации
 - Кнопка «Отмена»
 - 1) Данные не сохраняются
 - 2) Система возвращает пользователя к предыдущему экрану

- Кнопка «ОК»
 - 1) Проверка номера введённого ПИН-кода
 - 2) Система переводит пользователя на следующий экран

Важные процессы управления качеством проекта, о которых следует знать

Первый важный шаг в разработке высококачественных системных требований для группы системного провайдера и исследовательской группы — внимательно прочитать и изучить Протокол исследования и другие документы, описывающие исследование и стандарты спонсора. После чего основные члены группы системного провайдера должны провести подробное и продуманное обсуждение предъявляемых требований, чтобы определить, как именно группе клинических исследований необходимо, чтобы программа работала, чтобы поддерживать этот Протокол, качество и целостность данных, клинические сайты и персонал [13, 14]. Отдельно следует напомнить участникам, что нужно найти компромисс между сроками, стоимостью и качеством. Чем больше функций будет включено в систему, тем больше времени займет её разработка и проверка, и тем дороже это обойдется.

Обсуждение требований должно сопровождаться созданием первого чернового варианта документа системных требований. Этот первый черновик должен быть внимательно прочитан членами обеих групп. Ещё одно совместное обсуждение должно повлечь за собой решение вопроса, где и каким образом группа клинического исследования будет обеспечивать обратную связь и запросы о внесении изменений [6, 13, 14]. Процедура сбора требований, как правило, требует, по крайней мере, две ревизии первого черновика, чтобы создать окончательную документацию системных требований. Если рамки утверждённой системы требований превышают первоначальные предположения при планировании проекта, в план проекта могут быть внесены изменения, которые могут повлиять на график и бюджет исследования. Системный провайдер должен обеспечить контроль качества при производстве документации системных требований таким образом, чтобы независимая сторона ознакомилась с финальной версией документации системных требований прежде, чем её представят группе клинического исследования.

Дизайн (проект) системы

Что это такое? Этап проектирования включает в себя разработку проектной документации системы, обеспечивающей полное, чёткое, подробное техническое описание, каким образом система ePRO будет разрабатываться с учётом удовлетворения

потребностей Протокола и пользователей. Проектная документация необходима, чтобы объяснить разработчику, что должно быть включено, а что выходит за рамки системы. Эта документация должна охватывать как контекст, в котором система будет работать, так и детали, необходимые, чтобы освободить разработчика от неопределённости в вопросе построения системы [13, 14].

Системный провайдер ePRO отвечает за разработку большей части этой документации. Спонсору или CRO, ответственным за «подгрузку» данных ePRO в базу данных исследования, может потребоваться предоставить спецификации для переноса файлов данных системному провайдеру ePRO, чтобы их можно было использовать при создании/настройке модуля передачи данных.

Специалисты по разработке программного обеспечения используют проектную документацию системы, чтобы писать программы и/или задавать значения технических параметров для системы ePRO. Некоторые системы ePRO используют базовую систему, настроенную для использования в конкретном исследовании с применением параметров и настроек, с минимальными изменениями исходного кода в дополнение к базовой системе.

Первоначальная концепция требований к жизнеспособной системе может пересматриваться, как только система будет создана и подвергнута модульным тестам. После создания системы, авторам проектной документации необходимо эффективно общаться с группой клинического исследования и группой разработчиков программного обеспечения, чтобы гарантировать, что проект системы, введённый в эксплуатацию, по-прежнему соответствует требованиям пользователей [17].

Проектная документация

Документация по системному дизайну должна описывать три основные функции, которые соответствуют непосредственно следующим модулям в типичной системе ePRO:

- сбор и хранение данных (сбор данных по любой из нескольких технологий, которые будут храниться на сервере, контролируемом провайдером системы ePRO);
- веб-портал и оповещения (позволяющие исходным данным на сервере отображаться, отчётам создаваться, а предупреждениям — запускаться и рассылаться пользователям);
- трансфер данных (преобразование сохранённых данных в файлы передачи для отправки спонсору или его CRO).

Другие названия для этой документации. Она также может быть известна как проектная спецификация программного обеспечения, техническая спецификация проекта или документация по спецификации системы.

Роль группы клинических исследований. Роль группы клинических исследований складывается из проведения проверки и принятия проектной документации системы. Эта роль должна быть определена в начале жизненного цикла проекта, чтобы избежать непредвиденных задержек в развитии системы. Группа клинических исследований должна стремиться прояснить области, где системный провайдер неправильно истолковал требования в проектной документации или представил неполное описание решения. В тех случаях, когда документацию трудно интерпретировать, группа клинических исследований должна обратиться за помощью к техническим экспертам по данному вопросу, а не предполагать, что спецификация полностью и достоверно соответствует требованиям.

Почему это важно? Проектная документация системы является мостом между требованиями системы и разработчиками программного обеспечения. Создание этой документации является важным шагом для того, чтобы усилия по разработке программного обеспечения привели к решению, устраивающему пользователей системы и совместимому с клиническим протоколом [6, 13, 14].

Минимальное содержание. Документация должна охватывать все три модуля, описанных выше, в идеале с наличием отдельного документа или раздела для каждого из них. В случаях, когда используется базовая система, и технические характеристики базовой системы не меняются, системный провайдер ePRO должен сделать соответствующие части этой документации доступными для группы клинического исследования.

Раздел документации о проектировании модуля «системы сбора данных» необходим, чтобы упорядочить детали, которые ещё не вошли в документацию системных требований по таким темам, как:

- алгоритм обработки экранов;
- алгоритм контроля входа пользователя в систему и контроль процессов перехода между экранами / скриптами;
- редакционные проверки, которые препятствуют введению неверных данных или команд;
- оповещения пользователей;
- прочие алгоритмы работы с ошибками.

Кроме того, проектная документация должна содержать следующие детали, которые не вошли в документацию системных требований:

- 1) детали процессов отправки данных на сервер, алгоритмы и методы, доступные для отправки данных, в том числе процессов управления для обеспечения того, чтобы отправляемые данные являлись полными, точными и не дублировались;
- 2) детали процессов обработки частичного/неполного сохранения данных на сервер (например, в случае «отката системы»).

Раздел «веб-портал» должен освещать следующие вопросы документации, которые ещё не вошли в документацию системных требований проекта:

- вход в систему и пароль безопасности;
- группы безопасности и их влияние на доступ к данным;
- отображение клинических данных, в том числе необработанных данных, метаданных и аудита;
- отображение отчётов и переходов между ними, стандартного и кастомизированных (разработанных по запросу пользователя);
- оповещения для пользователей и логика их преключений;
- другие процессы (такие как автоматизация логики устройства или запросов на изменение данных).

Раздел «передача данных» должен содержать следующие детали, которые не вошли в документацию системных требований:

- синхронизация, частота и способ передачи данных;
- множество элементов данных для передачи и их формата;
- процесс преобразования данных в формате, определённом спонсором или его CRO;
- процессы управления для обеспечения того, чтобы переданные данные являлись полными, точными и не дублировались.

Важные процессы управления качеством проекта, которые следует иметь в виду

Системный провайдер должен обеспечивать контроль качества при производстве проектной документации системы таким образом, чтобы независимая сторона, знающая системные требования, предоставила отзыв на документацию, прежде чем документация будет предоставлена группе клинического исследования.

Кодирование/адаптация/разработка программного обеспечения

Что это такое? Процесс кодирования/адаптации/разработки ПО представляет собой процесс написания кода ПО на языке программирования или сборки и настройки модулей кода, которые уже были разработаны для удовлетворения потребностей данного конкретного исследования.

Почему это важно? Эти процессы являются блоками создания компьютерной системы, которая будет использоваться в клиническом исследовании. Правильное исполнение и документирование процесса поможет в создании систем, которые относительно просты в обслуживании и легко могут пройти проверку третьими лицами [12].

Вовлечённость группы клинического исследования. Отсутствует. Группа клинического исследо-

вания в процессе не участвует, поскольку процесс носит сугубо технический характер. Как правило, группа клинического исследования общается с разработчиками через менеджера и/или аналитика проекта, который гарантирует, что проектные требования правильно переведены на язык спецификаций для разработчиков.

Важные моменты управления качеством проекта, о которых следует знать. Как и в любом процессе, цель заключается в выявлении каких-либо проблем или дефектов как можно раньше и обеспечении их скорейшего решения [11]. В разработке должна присутствовать проверка выполненного кода, которая может быть выполнена одним или несколькими техническими экспертами, которые понимают языки и стандарты используемого метода разработки (так называемый «просмотр кода»). До введения нового фрагмента кода в общую структуру проекта исследования, разработчику (или его коллегам) следует проверить отдельные модули на предмет того, выполняются ли они должным образом и отвечают ли требованиям проекта, ограничениям и изначальным предположениям (так называемый «юнит-тестинг» или «тестирование модулей») [18].

Обнаружение каких-либо ошибок, вызванных качеством процесса при разработке программного обеспечения, должно быть задокументировано с указанием информации об условиях, при которых ошибки произошли, для возможности их последующего воспроизведения. См. Приложение 2, Описание процесса обеспечения качества, в Дополнительных материалах на <http://dx.doi.org/10.1016/j.jval.2013.04.002>. Ошибкам должен быть присвоен определённый статус (например, открыт/назначен/исправлен/закрит), чтобы они могли быть исправлены в системе и проверены второй стороной, прежде чем они могут считаться решёнными. Пока система находится в разработке, должна существовать возможность проследить за создаваемым кодом или модулями, использованными в отдельных элементах дизайна.

Тестирование ПО системным провайдером

Что это такое? Тестирование по всем пунктам, описанным в документации системных требований, имеет решающее значение для того, чтобы убедиться, что новая система соответствует всем ранее оговорённым требованиям. План тестирования описывает такую стратегию тестирования окружающей среды и подходы, которые имитировали бы реальные условия. Он также содержит полный набор тестов для покрытия всех потребностей системы [7, 13].

Вовлечённость группы клинического исследования. Отсутствует. Группа клинического исследования не участвует в процессе тестирования; это ответственность системного провайдера.

Почему это важно? Программы, необходимые для обеспечения систем для клинических исследований, состоят из многих тысяч комплексно взаимосвязанных строк кода. Даже при участии опытных, преданных делу программистов, которые осуществляют высококачественный процесс кодирования, сложность большинства систем для клинических исследований практически гарантирует, что ранние версии системы не будут работать в полном соответствии с документацией по системным требованиям. Каждая часть системы должна быть тщательно проверена.

Тестирование системы осуществляется после составления детального плана тестирования и самих тестов. Тщательно выполненный план тестирования гарантирует, что система в течение клинических исследований будет работать как положено [7, 13].

Минимальное содержание. Наиболее важными разделами плана тестирования качества являются стратегия тестирования и примеры тестов.

- *Стратегия тестирования.* В этом разделе, как правило, разработанном лидерами команды тестирования, должна быть задокументирована стратегия тестирования системы в условиях, максимально приближенных к реальному исследованию. Например, если исследование с зачислением пациентов в 10 странах продлится 12 месяцев с 2-недельным периодом подготовки и 9-месячным периодом лечения, стратегия тестирования должна описать, как будет создана тестовая среда, как данные будут собраны и проверены на этапе тестирования, чтобы симулировать эти условия, и как обеспечить точность и высокий уровень уверенности, что система будет работать надёжно в ходе исследования. Кроме того, стратегия тестирования должна документировать метод тестирования, чтобы указать, как будут учтены все системные требования.
- *Примеры тестов.* Также известные как примеры тестовых скриптов, они являются основой стратегии тестирования. Они помогают определить, совпадает ли именно эта система с заявленными требованиями. Каждый тест содержит пошаговые инструкции для инженера-тестировщика о том, как ввести набор предопределённых данных, а также подробное описание ожидаемых результатов. При выполнении каждого теста, инженер-тестировщик будет должен чётко указать, совпали ли фактические результаты с ожидаемыми и, таким образом, пройден тест или «провален». Если «провален», инженер будет подробно документировать отклонение от ожидаемых результатов в выпускном отчёте, чтобы программист смог определить причину сбоя, решить её, и задокументировать это решение.

Важные варианты тестирования, о которых нужно знать

Необходимо использовать несколько вариантов тестирования. Это сведёт к минимуму риск системных сбоев и ошибок файлов данных при использовании системы в реальных условиях. Процесс тестирования должен включать следующие проверки каждого из требований:

- *положительное тестирование.* Широкий спектр допустимых значений и параметров, проводится, чтобы гарантировать, что допустимые значения будут приняты системой;
- *границные испытания.* Записи делаются только в пределах граничных значений (например, при ограничении по возрасту) и сразу за пределами граничных значений, чтобы подтвердить, что в поле «редакторские проверки на соответствие» показаны правильные результаты;
- *отрицательное тестирование.* Делаются допущения и/или неправильный ввод данных, чтобы удостовериться, что все возможные ошибки обрабатываются корректно;
- *нагрузочное тестирование.* Большой набор предопределённых достоверных данных вводится и проверяется как на локальной системе и на удалённых системах (например, хранилищах данных/серверах), куда данные будут передаваться в течение всего процесса, путём сравнения с предопределёнными ожидаемыми наборами данных. Размер набора данных должен быть достаточно большим для того, чтобы обеспечить всем связанным системам — локальной системе, удалённым системам, системе телекоммуникаций, и т.д. — возможность легко вместить существенно превосходящий ожидаемый объём данных для изучения, и всё равно выдавать точные результаты. Нагрузочное тестирование является жизненно важным для свидетельства того, что системы будут функционировать должным образом и в клиническом исследовании;
- *регрессивное тестирование.* Это очень важно после решения программистом одного или нескольких вопросов. Целью регрессивного тестирования является подтверждение, что изменение кода не повредило другим частям системы, которые были разработаны ранее. Регрессивное тестирование необходимо, потому что изменения в коде часто имеют непредвиденные последствия для других областей программного обеспечения. В случаях существенных изменений в коде, вся система должна быть проверена повторно для того, чтобы она продолжала работать на должном уровне. Когда изменения кода очень незначительные, например, исправления орфографи-

ческих ошибок или других мелких ошибок, регрессивное тестирование может проводиться в меньших масштабах. Группа тестирования должна очень хорошо просчитать уровень риска, возникающий вместе с изменением кода, и использовать стратегию регрессивного тестирования, что позволит значительно минимизировать риск.

Важные процессы управления качеством, о которых нужно знать. Тестовые стратегии, планы и тестовые случаи должны пройти несколько проверок качества управления и несколько итераций (повторений) в команде системного провайдера. При выполнении тестов часто обнаруживаются вопросы или ошибки, которые должны быть тщательно задокументированы и отслеживаются системой тестеров. Программисты будут создавать и документировать новую версию системы, которая содержит изменения, внесённые, чтобы решить ошибки и проблемы системы. Потом будет создана новая стратегия тестирования для каждой новой версии системы. Эта новая стратегия тестирования должна учитывать риск того, как отразятся изменения программного обеспечения на ранее выполненных системных функциях. На основе этой оценки рисков, стратегия тестирования должна будет учесть все тестовые случаи, которые будут перезапущены, чтобы гарантировать, что все ошибки были решены, и для того, чтобы регрессивное тестирование было проведено адекватно, необходимо в полной мере учитывать риск изменения программного обеспечения и обеспечения надёжной работы системы в условиях исследования [7, 13].

Если качество процессов было соблюдено системным провайдером верно, потребуется от двух до трёх итераций стратегии тестирования и тестовых случаев. Заключение — это успешное прохождение теста на приемлемость для пользователя группой исследования, и получение системы, работающей так, как предполагается в течение клинического исследования.

Трассировка (traceability)

Что это такое? Трассировка (прослеживаемость) играет роль контроля качества путём установления, что программное обеспечение удовлетворяет потребностям пользователя. Это гарантирует, что все элементы в документации системных требований должным образом «трассируются» в другие процессы и документы проверки критически важных систем.

Матрица трассировки показывает, что каждый пункт в системных требованиях был учтён при разработке документа, а также в случаях тестирования или скриптах. Она может быть использована для отслеживания:

- 1) требований дизайна (пользователя) к функции системы;
- 2) требований программного обеспечения (функциональных), определённых для удовлетворения требований пользователя;
- 3) тестовые примеры, которые используются для проверки того, что конечный продукт отвечает требованиям программного обеспечения и пользовательским требованиям.

Тестовые матрицы могут принимать различные формы. Например, трассировка может быть проведена с юнит-тестирования требований дизайна, таких как тестирование фрагмента кода разработчиком, чтобы увидеть, работает ли он в изоляции. Другой пример — трассировка с тестового случая (формальное тестирование конкретной спецификации программного обеспечения) на требование программного обеспечения.

Почему это важно? Матрица трассировки гарантирует, что каждое системное требование было разработано и закодировано в программном обеспечении и включено в тестовую программу. Матрица трассировки также гарантирует, что все требования были апробированы в рамках системного и пользовательского тестирования [13, 14, 17]. Обеспечение прозрачности трассировки в течение всего рабочего процесса необходимо, чтобы все элементы дизайна были не только надлежащим образом протестированы, но и позволяли проводить анализ в дальнейшем, когда инспектор будет проводить валидацию путём тестирования записей по критическим функциям дизайна.

Чаще всего не существует прямой корреляции между проектными требованиями, требованиями к программному обеспечению и тестами. Например, на одно проектное требование может приходиться несколько функциональных требований, чтобы охватить всю функцию; на одно функциональное требование может потребоваться несколько тестов, чтобы убедиться, что всё работает правильно. Таким образом, матрица трассировки документирует эти зависимости между функциональными требованиями пользователей и техническими требованиями программного обеспечения.

Минимальное содержание. За счёт матрицы трассировки каждое требование системы к элементу конструкции и тестовым примерам является минимальным. Матрица может быть самостоятельным документом или кратко излагаться в утверждённом испытательном отчёте, показывающем, каким образом тесты проверяют все требования к конструкции.

Вовлечённость группы клинического исследования. Отсутствует. Трассировка является обязанностью системного провайдера. Спонсор может проверить и подтвердить эту работу в рамках аудитов.

Пример исследования - спецификация проекта ePRO к требованиям матрицы трассировки											
Список требований	Спецификации проекта										Покрытие требований
	Спец-1	Спец-2	Спец-3	Спец-4	Спец-5	Спец-6	Спец-7	Спец-8	Спец-9	Спец-10	
Треб-1	•	•	•								100%
Треб-2А				•							100%
Треб-2В					•	•					50%
Треб-2С							•				100%
Треб-3								•			100%
Треб-4А									•		100%
Треб-4В									•		100%
Треб-5										•	75%

Рис. 1. Пример матрицы трассировки *

Примечание: в некоторых случаях, одна спецификация полностью покрывает все нужды. Однако временами для покрытия требований требуется исполнение нескольких спецификаций. (См. Спец-1 в таблице). Требования 2В и 5 не удовлетворены полностью. Будет необходимо вернуться на шаг назад и пересмотреть документацию, переделав её таким образом, чтобы полностью соответствовать требованиям спонсора.

Важные процессы управления качеством, о которых нужно знать. Тщательная подготовка матрицы трассировки на этом этапе может помочь предотвратить ошибки в проектировании, исправление которых будет более сложным и дорогостоящим в течение жизненного цикла системы. Как и все процессы и системы, подлежащие контролю качества, матрица трассировки должна быть изучена сотрудниками системного провайдера, независимыми от команды разработчиков.

Тестирование на приемлемость для пользователя (user acceptance testing, UAT)

Что это такое? UAT — это процесс, при котором группа клинического исследования определяет, соответствует ли система ожиданиям и работает ли в соответствии с документацией системных требований. При существовании несоответствия между ожиданиями группы клинического исследования и системы, они будут выявлены на этом этапе. UAT можно начинать после того, как системный провайдер представит письменное подтверждение выполнения своих обязательств по валидации системы. Следует отметить, что UAT — это не полная перепроверка системы, предпринимаемая группой клинического исследования. Скорее, это сфокусированный, риск-ориентированный подход к тестированию, который позволяет группе клинического исследования определить, соответствует ли система ключевым требованиям к системе (которые отражает Протокол).

Другие названия для этой документации. UAT также известно как тестирование пользователями сайта.

Вовлечённость группы клинического исследования. Группа клинического исследования должна принимать активное участие в разработке стратегии для UAT. Это не может находиться только в ответственности системного провайдера или третьей стороны.

Авторы Протокола, формулировавшие требования к системе, вначале должны проверить и подтвердить соблюдение этих требований [6].

Как правило, в группе клинического исследования присутствует несколько членов, участвующих в выполнении тестовых скриптов. Группа клинического исследования играет решающую роль в рассмотрении, готова ли система к выпуску, за счёт проверки расхождений, выявленных в UAT и способов их решения (их либо устраняют, либо соглашаются, что расхождения не требуют исправления). После того как группа клинического исследования определяет, что система готова к производству, она должна подготовить и подписать официальный документ, который определяет систему, как утверждённую для выполнения поставленных задач [6, 7].

Следует отметить, что UAT заметно отличается от юзабилити-тестирования. В отчёте рабочей группы ISPOR по ePRO от 2009 года [1], юзабилити-тестирование было определено как проверка, «могут ли респонденты из целевой группы населения использовать программное обеспечение и устройства соответствующим образом. Этот процесс включает в себя формальную документацию способности респондентов использовать навигацию электронной платформы, следовать инструкциям и отвечать на вопросы». Проще говоря, юзабилити-тестирование позволяет определить, могут ли субъекты провести компьютеризированную оценку как предполагалось?

Юзабилити-тестирование включает в себя участие представителей населения, вовлечённых в клиническое исследование. Например, в юзабилити-тестировании системы ePRO для педиатрического исследования, которое включает собственный отчёт ребенка, должны участвовать дети. UAT не задействует субъекты целевой популяции; оно проводится между группой клинического исследования и провайдером ePRO. Хотя и UAT, и юзабилити-тестирование необходимы и важны для общего успе-

ха исследования, UAT имеет решающее значение для определения, соответствует ли программное обеспечение требованиям документации для пользователя и письменным спецификациям.

Почему это важно? UAT является важным шагом для проверки того, что система была построена в соответствии с первоначальным документом о требованиях пользователя. Тщательные усилия по проведению UAT также предоставляют группе клинического исследования возможность увидеть, как система будет функционировать и самообучаться на деталях, необходимых для ответа на вопросы исследовательских сайтов и аудиторов, как только система будет запущена в работу. Кроме того, это помогает укрепить уверенность группы клинических исследований, что система будет надёжно функционировать в течение всего процесса.

После успешного завершения UAT даст ответы на три основных вопроса:

1. Была ли система спроектирована и построена в соответствии с первоначальными требованиями пользователя?
2. Соответствуют ли первоначальные требования всем ожиданиям команды клинического исследования?
3. Создан ли правильный дневник субъектов для использования в этом исследовании? [17].

Хотя у группы клинического исследования может возникнуть желание положиться на валидацию системы провайдером и пропустить UAT, это запрещено правилами Международной конференции по гармонизации технических требований к регистрации фармацевтической продукции применяемой для людей в надлежащей клинической практике (The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use Good Clinical Practice) [19], которые требуют, чтобы спонсор брал на себя полную ответственность за качество и целостность данных исследования. Спонсор несёт ответственность за внедрение и поддержание системы обеспечения качества и системы контроля качества с помощью письменных Стандартных Операционных Процедур (СОП), чтобы быть уверенным, что испытания проводятся и данные генерируются, документируются (записываются), и передаются в соответствии с Протоколом, надлежащей клинической практикой (GCP) и нормативными требованиями [19]. Увеличить процесс разработки на несколько недель — это небольшая цена за предотвращение потока неприятных находок во время проверки регуляторной инспекцией.

Минимальное содержание. Чтобы правильно выполнить UAT, группа клинического исследования документирует стратегию тестирования в тестовом плане UAT [20]. Эта стратегия будет направлять раз-

витие сценариев тестирования, задавать соответствующий набор тестеров, и определять тестовый период (в том числе повторное тестирование в случае необходимости). Это похоже на стратегию тестирования внутренней проверки системным провайдером, но не так обширно.

Тестовый план UAT должен касаться и тех областей, которые требуют тщательного тестирования и тех, которые требуют лёгкого тестирования, основанного на восприятии риска [12]. Например, требование о том, чтобы программное обеспечение вычисляло суммарный балл на основе ответов в новом дневника, где эта оценка используется для определения того, будет ли субъект рандомизирован в исследование, будет иметь большое значение; такая задача потребует тестовых примеров, определяющих условия как для успешной рандомизации, так и для невключения субъекта. Требование, чтобы будильник на устройстве звучал точно в 5:00 вечера, будет нуждаться только в беглом тестировании во время UAT, поскольку будильник, как правило, является стандартной функцией базового программного обеспечения.

Важные процессы управления качеством, на которые надо обратить внимание. Надёжность UAT можно повысить путём использования матрицы трассировки для уст тестов UAT к исходным требованиям. Матрица — это эффективный способ определить, все ли изначальные условия удовлетворены за счёт UAT. Кроме того, необходим независимый контроль качества тестовых скриптов, чтобы свести к минимуму риск использования скриптов, которые не могут выполнить функцию в соответствии с проектом или не достигнут своих конкретных целей тестирования [13].

UAT иногда приводит к тому, что группа клинического исследования понимает, что система функционирует как записано и спроектировано, но не удовлетворяет задокументированным требованиям или ожиданиям, включённым в Протокол. Если группа клинического исследования уверена в необходимости внесения изменений в то, как система в настоящее время функционирует, то команда обязана инициировать «изменения в масштабах и сфере деятельности» проекта. Требования переписывают, и повторяется весь процесс жизненного цикла системы, пока следующий UAT не покажет необходимого результата. Следует отметить, что группа клинического исследования может поднимать вопрос о пересмотре требований на таком позднем этапе только при наличии времени, бюджета и по-настоящему важного несоответствия, так как эти изменения могут повлиять на выбор первого субъекта/первую дату визита.

Качество нельзя внедрить в систему за счёт тестирования. UAT служит в качестве контрольного фильтра. Если процесс документирования требований

системным провайдером, проектирования системы или её создания в соответствии с проектной документацией нарушается, необходима обратная связь между спонсором и системным провайдером, чтобы, в первую очередь, ускорить процесс внесения улучшения качества. Такая обратная связь может быть реализована путём проведения встреч об «извлечённых уроках», и требует совместных усилий спонсора и провайдера по устранению негативных находок, обнаруженных до начала нового проекта.

Процесс должен быть организован так, чтобы получить формальное документирование усвоенных уроков и шагов, которые были предприняты по их решению. Такой подход поможет системному провайдеру сделать существенные улучшения в процессе валидации системы, которые должны привести к более гладкому проведению UAT в будущих проектах. Наконец, UAT системы ePRO не завершает процесс валидации системы. До момента, пока система ePRO не будет запущена на рабочем месте конечного пользователя, этот процесс не считается завершённым [7, 8, 13].

Менеджмент установки/конфигурации

Что это такое? Менеджмент установки/конфигурации представляет собой процесс, когда системный провайдер устанавливает полностью протестированное ПО, используя методы, описанные выше, на устройство ePRO. Он включает в себя локализацию системы для конечного пользователя, предоставляя ему инструкции по эксплуатации и план технической поддержки системы. Он также включает в себя базовые и специфические, связанные с исследованием, установки сервера. Чтобы убедиться, что правильная версия исследования будет применяться в правильной локации, требуется надёжный процесс управления конфигурацией/установкой. Это верно как для исследовательских серверов, так и для клиентских компонентов. После однократной установки на целевых средах, система может рассматриваться как «валидированная» [7].

Почему это важно? Поскольку многие клинические испытания являются глобальными, установка ePRO с правильной версией программного обеспечения и локальными настройками может быть сложной. Процесс настройки очень важен для того, чтобы «окончательная», «прошедшая тестирование» версия системы и настройки были правильно локализованы (например, версия для Пенджаба должна быть установлена на правильном языке (языках) и в часовом поясе, верном для Пенджаба, Индия). Важно также убедиться, что физические аксессуары, такие как источники питания, подходят для исследовательского сайта (места проведения исследования), и что материалы обучения или пользования предоставлены на соответствующем языке (языках).

Вовлечённость группы клинического исследования. Группа клинического исследования должна быть осведомлена о процессах провайдера ePRO для управления конфигурацией, установки и логистики поставки на многочисленные глобальные сайты. Вовлечение группы клинического исследования необходимо для того, чтобы на сайт поставлялся соответствующий локализованный продукт.

Важные моменты управления качеством, которые нужно иметь в виду. Установка ePRO провайдером и конфигурационная группа должны использовать тщательно разработанный контрольный список качества, чтобы управлять процессом установки программного обеспечения для каждого решения ePRO. Контрольный список содержит документацию для настроек каждого устройства, внутренних проверок качества настроек и отчётность персонала, ответственного за установку и конфигурацию.

И, наконец, рекомендуется контрольная выборка качественной статистики по устройствам для проведения валидации на выборке субъектов для статистического контроля качества, чтобы удостовериться, что устройства и аксессуары, предназначенные для конкретного региона, являются правильными. Конечный результат заключается в обеспечении уверенности, что группа клинического исследования установила систему ePRO для сбора данных от субъектов по всему миру, и она может пройти регуляторную проверку [7, 13].

Выведение из эксплуатации

Что это такое? Выведение из эксплуатации — это процесс, осуществляемый системным провайдером, когда система для ввода данных и сервисов выводится из эксплуатации или списывается по окончании клинического исследования. Процесс вывода из эксплуатации гарантирует, что все открытые вопросы сняты с рассмотрения и закрыты. Этот последний шаг в процессе валидации состоит из следующих этапов:

1. *Завершение работы с данными:* гарантия того, что все данные пациента, которые были собраны, загружены из устройств пациентов в центральную базу данных провайдера ePRO. Это сопровождается отключением дальнейшей загрузки данных и блокировки базы данных для скачивания в соответствии с потребностями группы исследования.

2. *Возвращение устройств* (конкретно — КПК): подсчёт возвращённых устройств, а также их очистка (как от основного ПО, так и от сопутствующих), переработка для последующих исследований или утилизация.

3. *Документация:* гарантии, что инвентаризация всех необходимых документов проверки и записи находятся в архивном хранилище поставщика.

4. *Уведомления:* уведомление всех внутренних и внешних сторон поддержки и отказ от услуг, которые более не потребуются.

Почему это важно? Вывод из эксплуатации гарантирует, что устройства, системы и услуги, настроенные для исследования, больше не находятся в использовании, что все собранные данные были переданы исследовательской группе, и остались только официальные копии.

Вовлечённость группы клинического исследования. Группа должна быть осведомлена о процессах вывода системы из эксплуатации провайдером ePRO. Это подтверждает конфиденциальность и позволяет избежать непреднамеренного или ошибочного сбора данных.

Важные процессы управления качеством, о которых нужно знать. Выводящая ePRO из эксплуатации команда провайдера должна использовать тщательно разработанный контрольный список, который обеспечивает документирование процесса на каждом устройстве, внутреннюю оценку качества процессов и отчётность сотрудников, исполняющих эти процессы [7, 13].

Выводы

Валидация электронных систем для сбора данных PRO является критически важным компонентом для получения разрешения регуляторных органов на проведение клинического исследования. В этом отчёте рассматриваются технические аспекты систем сбора данных ePRO и процесса их валидации, а также распределение обязанностей при совместной разработке системы группой клинического исследова-

ния и системным провайдером. В результате, отчёт должен улучшить понимание требований к проверке технологий спонсорами клинических исследований, чтобы обеспечить основу для сравнения различных поставщиков систем ePRO и их соответствующих предложений об услугах.

Благодарности

Авторы искренне благодарны следующим членам ISPOR за их вклад в разработку этих совместных надлежащих исследовательских практик, указанных в данном отчёте: *Katy Benjamin, Stephen Coons, Sonya Eremenco, Ancilla Fernandes, Alison Greene, Parul Gupta, Syed Umer Jan, Sandra Kane-Gill, Smita Kothari, Tamar Lasky, Yvonne Lis, Damian McEntegart, Wilhelm Muhlhause, Hannah O’Gorman, Raymond Panas, Yong Joo Rhee, Bob Smith, Jun Su, Etta J. Vinik, and Shaowei Wan.*

Источник финансовой поддержки

В поддержку этого исследования не было вложено никаких средств.

Дополнительные материалы

Дополнительные материалы к этой статье можно найти в онлайн-варианте по гиперссылке <http://dx.doi.org/10.1016/j.jval.2013.04.002>

Или, при необходимости строгой копии статьи: www.valueinhealthjournal.com/issues (выберите том, тему и статью).

Литература

1. Coons S.J., Gwaltney C.J., Hays R.D., et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force Report. // Value Health 2009;12:419-29. Available from: http://www.ispor.org/workpaper/patient_reported_outcomes/Coons.pdf. [Accessed March 26, 2013].
2. US Food and Drug Administration. Guidance for industry: patient reported outcome measures: use in medical product development to support labeling claims. December 2009. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>. [Accessed February 19, 2011].
3. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health related quality of life (HRQL) measures in the evaluation of medicinal products. 2005. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003637.pdf. [Accessed February 22, 2011].
4. Paty J., Stokes T. Electronic diaries, part 1: what is a subject diary, and how do regulations apply. // Appl Clin Trials 2002. Available from: <http://www.appliedclinicaltrials.com/appliedclinicaltrials/article/articleDetail.jsp?id=83521>. [Accessed March 23, 2011].
5. Hufford M.R., Stokes T.E., Paty J.A. Collecting reliable and valid real-time patient experience data. // Drug Inf J 2001;35:755–65.
6. Paty J., Stokes T. Electronic diaries, part 2: the role of the clinical protocol in developing and implementing electronic diaries. Appl Clin Trials 2003. Available from: <http://www.appliedclinicaltrials.com/appliedclinicaltrials/article/articleDetail.jsp?id=90715>. [Accessed March 23, 2011].
7. Chamberlain R. Computer Systems Validation for the Pharmaceutical and Medical Device Industries. (2nd ed.). Libertyville, IL: Alaren Press, 1994.
8. Gogates G.D. Software validation in accredited laboratories: a practical guide. June 2010. Available from: ftp://ftp.fasor.com/pub/iso25/validation/adequate_for_use.pdf. [Accessed July 5, 2012].
9. US Food and Drug Administration. Guideline on general principles of process validation. 1987. Available from: <http://www.fda.gov/MedicalDevices/Device-RegulationandGuidance/PostmarketRequirements/QualitySystemsRegulations/MedicalDeviceQualitySystemsManual/ucm122439.htm>. [Accessed September 3, 2011].
10. US Food and Drug Administration. Guidance for industry process validation: general principles and practices. 2011. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070336.pdf>. [Accessed September 3, 2011].
11. European Commission Health and Consumers Directorate-General. The rules governing medicinal products in the European Union, volume 4: good manufacturing practice: medicinal products for human and veterinary use, annex 11: computerised systems. June 2011. Available from: http://ec.europa.eu/health/files/eudralex/vol-4/annex11_01-2011_en.pdf. [Accessed September 1, 2011].

12. U.S. Food and Drug Administration. General principles of software validation: final guidance for industry and FDA staff. January 2002. Available from: <http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126955.pdf>. [Accessed August 12, 2011].
13. Stokes T. The Survive and Thrive Guide to Computer Validation. Buffalo Grove, IL: Interpharm Press, Inc., 1998.
14. Stokes T., Branning R.C., Chapman K.G., et al. Good Computer Validation Practices: Common Sense Implementation. Buffalo Grove, IL: Interpharm Press, Inc., 1998.
15. US Food & Drug Administration. Guidance for industry part 11, electronic records; electronic signatures — scope and application. August 2003. Available from: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm072322.pdf>. [Accessed August 15, 2011].
16. Cardie M.L., Tucker N.G., Weiss L.M. Computer system validation. GMP Rev 2005;4:20—2. Fig. 1 — Sample of traceability matrix. 488 // Value Health 16 (2013) 480 — 489.
17. Stokes T., Paty J. Electronic diaries, part 3: developing and validating electronic diaries: roles for the technical and clinical teams. // Appl Clin Trials 2003;6:68—78.
18. PIC/S. Good practices for computerised systems in regulated “GXP” environments, section 13 “testing” report PI 011-3, Pharmaceutical Inspection Convention, Geneva. September 2007. Available from: <http://www.picscheme.org/publication.php?id=8>. [Accessed August 12, 2012].
19. European Medicines Agency (EMA) ICH Topic E 6 (R1) Guideline for good clinical practice. July 2002. Available from: <http://ichgcp.net/pdf/ich-gcp-en.pdf>. [Accessed September 21, 2012].
20. Atkins T. User acceptance testing: finally some validation? Silverpath Technologies. 2009. Available from: <http://silverpath.com/resources/Silverpath-UserAcceptanceTestingWhitepaper-090203.pdf>. [Accessed March 26, 2013].

Принципы надлежащей практики для применения аналитического моделирования в оценке медицинских технологий: доклад Рабочей группы ISPOR по надлежащей практике исследований — исследования по моделированию

Milton C. Weinstein¹, Bernie O'Brien², John Hornberger³, Joseph Jackson⁴, Magnus Johannesson⁵, Chris McCabe⁶, Bryan R. Luce⁷

¹ — Center for Risk Analysis, Harvard School of Public Health, Boston, MA, USA, and Innovus Research, Inc., Medford, MA, USA

² — Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada

³ — Acumen, LLC, and Stanford University School of Medicine, Stanford, CA, USA

⁴ — Pharmaceutical Research Institute, Bristol-Myers Squibb, Princeton, NJ, USA

⁵ — Centre for Health Economics, Stockholm School of Economics, Stockholm, Sweden

⁶ — Trent Institute for Health Services Research, University of Sheffield, Sheffield, UK

⁷ — MEDTAP International, Bethesda, MD, USA

Переводчики: Павлыш Андрей Владиславович — к.м.н., докторант Первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Рецензенты: Колбин Алексей Сергеевич — д.м.н., профессор, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, г. Санкт-Петербург, Российская Федерация

Резюме. Цели: Математическое моделирование широко используется при экономической оценке фармацевтических и других технологий здравоохранения. Пользователи моделей, как в правительстве, так и в частном секторе должны быть в состоянии оценить качество моделей в соответствии с научными критериями надлежащей практики. В настоящем докладе описываются заключения Рабочей группы, созданной для разработки основных принципов проведения и представления исследований с использованием моделирования. **Методы:** Рабочая группа была назначена по совету и с согласия Совета директоров ISPOR. Членами группы выступили опытные разработчики или пользователи моделей из нескольких стран Северной Америки и Европы, работающие в сфере науки и промышленности. Рабочая группа встречалась трижды, обменивалась идеями в процессе многократной переписки по электронной почте, а также консультировалась с основной группой внешних экспертов и с экспертами из числа членов ISPOR. **Результаты:** Критерии оценки качества моделей были распределены в соответствии с тремя составляющими, относящимися к структуре моделей, используемым входным и выходным данным, и валидации моделей. При рассмотрении этих составляющих было затронуто несколько основных тем. В частности, тот факт, что модели и их результаты должны быть представлены в качестве вспомогательной информации, облегчающей процесс принятия решений, но не в качестве утверждений, претендующих на научную истинность.

В связи с этим, нецелесообразно требовать валидации моделей до их использования. Тем не менее, предположения, лежащие в основе моделей и касающиеся структуры причинно-следственных связей и значений параметров, должны постоянно сравниваться с имеющимися данными, и при необходимости в модели должны быть внесены соответствующие коррективы. Предположения о структуре моделей, а также оценки параметров должны быть представлены чётко и ясно. Проведение анализа чувствительности должно обеспечить пользователям моделей возможность оценить условные связи между входными и выходными переменными моделей. *Выводы:* Оценки, основанные на использовании моделей, являются значимым ресурсом для лиц, принимающих решения в области здравоохранения. Разработчики моделей обязаны проводить процесс моделирования в соответствии с лучшими существующими стандартами качества и представлять конечные результаты, должным образом указывая предположения и ограничения моделей и разъясняя то, что полученные выводы обусловлены вышеуказанными предположениями и исходными данными, на которых строились модели.

Ключевые слова: математическое моделирование, оценка технологий здравоохранения, принятие решений, валидация моделей, анализ чувствительности

Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices — Modeling Studies

Abstract. Objectives: Mathematical modeling is used widely in economic evaluations of pharmaceuticals and other healthcare technologies. Users of models in government and the private sector need to be able to evaluate the quality of models according to scientific criteria of good practice. This report describes the consensus of a task force convened to provide modelers with guidelines for conducting and reporting modeling studies. **Methods:** The task force was appointed with the advice and consent of the Board of Directors of ISPOR. Members were experienced developers or users of models, worked in academia and industry, and came from several countries in North America and Europe. The task force met on three occasions, conducted frequent correspondence and exchanges of drafts by electronic mail, and solicited comments on three drafts from a core group of external reviewers and more broadly from the membership of ISPOR. **Results:** Criteria for assessing the quality of models fell into three areas: model structure, data used as inputs to models, and model validation. Several major themes cut across these areas. Models and their results should be represented as aids to decision making, not as statements of scientific fact; therefore, it is inappropriate to demand that models be validated prospectively before use. However, model assumptions regarding causal structure and parameter estimates should be continually assessed against data, and models should be revised accordingly. Structural assumptions and parameter estimates should be reported clearly and explicitly, and opportunities for users to appreciate the conditional relationship between inputs and outputs should be provided through sensitivity analyses. **Conclusions:** Model-based evaluations are a valuable resource for health-care decision makers. It is the responsibility of model developers to conduct modeling studies according to the best practicable standards of quality and to communicate results with adequate disclosure of assumptions and with the caveat that conclusions are conditional upon the assumptions and data on which the model is built.

Keywords: mathematical modeling, health technology assessment, decision-making, model validation, sensitivity analysis

Опубликовано: Milton C. Weinstein, Bernie O'Brien, John Hornberger, Joseph Jackson, Magnus Johannesson, Chris McCabe, Bryan R. Luce. Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices — Modeling Studies. // Value Health 2003; 6:9-17.

Введение

Математическое моделирование широко используется в экономических оценках фармацевтических и других технологий здравоохранения. Цель моделирования состоит в структурировании доказательной базы клинических и экономических результатов в той форме, которая может помочь в процессе принятия решений о клинической практике и распределении ресурсов здравоохранения.

Данные о последствиях для здоровья и затратах на лечение собираются из различных источников, в том числе используются данные клинических и эмпирических исследований, базы данных по страховым случаям, реестры историй болезней, статистические данные по здравоохранению и исследования предпочтений. Модель является логической математической основой, которая позволяет объединить качественные и количественные данные и соединить

их с результатами, представляющими интерес для лиц, принимающих решения в области здравоохранения. При принятии решений о распределении ресурсов конечным результатом модели часто является оценка стоимости дополнительного года жизни с поправкой на качество (QALY) или иная мера соотношения затрат и качества.

Несмотря на то, что основными при исследовании действенности остаются данные из рандомизированных клинических исследований (РКИ), взятые отдельно, они могут привести к неправильным выводам, в случае, если конечные критерии оценки не преобразованы в показатели, которые являются значимыми для пациентов, провайдеров, страховых агентов и широкой общественности. Например, предположим, что согласно результатам РКИ, лечение снижает риск редкого осложнения хронического заболевания на 50%. Далее, предположим, что другое исследование показывает, что иная методика лечения

снижает риск другого, более распространённого осложнения, на 10%. Последнее лечение может быть более эффективным и экономически целесообразным, чем первое, но простого сравнения результатов исследований будет недостаточно. Использование модели может помочь при представлении полученных выводов лицам, принимающим решения. Сравнение двух стратегий лечения будет зависеть от синтеза данных о частоте возникновения рассматриваемых осложнений в целевой группе населения, сокращении относительного риска, вызванного предлагаемым лечением, выживаемости и качестве жизни с учётом возникших осложнений и без них, а также расходов на лечение и необходимую медицинскую помощь для диагностики и купирования данных осложнений.

Ценность модели заключается не только в результатах, которые она предоставляет, но и в её способности выявлять логические связи между входными параметрами (т.е. данными и предположениями) и выходами в виде оцененных исходов и затрат. По этой причине модель не должна быть своего рода «чёрным ящиком» для конечного пользователя. Наоборот, она должна быть как можно более прозрачной, для того чтобы логику, лежащую в основе модели и её результатов, можно было бы понять на интуитивном уровне. По этой же причине результаты моделирования никогда не следует представлять в виде точечных оценок или в виде безусловных утверждений о ответственности или затратах. Необходимо указывать на факт того, что полученные результаты действительны только при условии использования допустимых предположений и конкретных входных данных, и они должны включать анализ чувствительности для изучения влияния альтернативных данных и предположений на результаты модели.

Цель этой статьи заключается в представлении общей позиции Рабочей группы ISPOR по надлежащей практике исследований на основе моделирования. Как и сами модели, эта позиция, являясь результатом обсуждений Рабочей группы на настоящий момент, может быть изменена благодаря появлению новых технологий для моделирования и распространению таких принципиально новых технологий здравоохранения и окружающей среды как геномная резистентность или резистентность микробов к лекарственным препаратам.

Предыстория создания Рабочей группы

Председатель Рабочей группы ISPOR по надлежащей практике исследований с использованием моделирования, *Milton C. Weinstein*, был назначен в 2000 году Председателем Совета по науке в здравоохранении в ISPOR, *Bryan R. Luce*. Члены Рабочей группы были приглашены для участия Председателем, по рекомендации и с согласия Совета директоров ISPOR. Мы искали людей, которые бы обладали опытом в

разработке или использовании фармакоэкономических моделей; ведущих научных специалистов в данной области; людей, работающих в академической и промышленной сферах и в качестве советников в государственных учреждениях; и которые представляли бы различные страны и системы здравоохранения. Была также создана референтная группа из членов ISPOR, которые предоставляли комментарии и замечания по работе. Рабочая группа провела своё первое заседание на ежегодном Северо-Американском Научном Конгрессе ISPOR в Арлингтоне, штат Вирджиния, в мае 2000 г. В последующие месяцы для обмена идеями членами Рабочей группы широко использовалась электронная почта. Председателем был подготовлен проект доклада, который был разослан членам Рабочей группы для рассмотрения и внесения возможных изменений и комментариев. Доработанный проект был разослан членам референтной группы, и после получения её замечаний был сформирован новый проект доклада. Резюме данного проекта доклада было представлено на пленарном заседании ежегодного Северо-Американского Научного Конгресса ISPOR в Арлингтоне, штат Вирджиния, в мае 2001 года. Комментарии от слушателей были включены в переработанный проект, который был размещён на веб-сайте ISPOR для общих замечаний. Следующий вариант проекта доклада был представлен на ежегодной европейской научной конференции ISPOR в Каннах, Франция, в ноябре 2001 года, и исправленный проект был опубликован для дальнейших комментариев на веб-сайте ISPOR. Данный доклад отражает все внесённые замечания и комментарии на всех перечисленных этапах.

Определение Модели

Национальный исследовательский совет в своем докладе по вопросам использования микромоделирования в задачах социальной политики предложил следующее определение имитационной модели: «... воспроизводимая, подтверждённая последовательность вычислений, используемая для создания оценок интересующих величин...» [1]. Под моделью оценки технологий здравоохранения мы понимаем аналитическую методологию, которая учитывает события во времени и в различных популяциях, которая основана на данных, полученных из первичных и/или вторичных источников, и цель которой заключается в оценке результатов вмешательства на последствия для здоровья и соответствующие затраты.

В рамках нашего рабочего определения будем считать, что модели экономической эффективности предназначены для помощи в процессе принятия решений. Это означает, что их целью является не делать безоговорочные утверждения о последствиях того или иного медицинского вмешательства, но выявлять связь между предположениями модели и результата-

ми. Эти предположения включают в себя следующие аспекты: структурные предположения о причинно-следственных связях между переменными; количественные параметры, такие как частота выявления новых случаев заболевания и распространенность заболевания; эффективность и действенность стратегии лечения; выживаемость; шкала оценки состояния здоровья; интенсивность использования; затраты; а также оценочные суждения, такие как характер последствий, имеющих значение для лиц, принимающих решения. Качественное исследование, основанное на модели, представляет выводы, обязательно учитывающие вышеперечисленные предположения, уделяя внимание их понятности и прозрачности.

Оценка модели

Модели следует использовать только после тщательного тестирования, чтобы гарантировать, что математические расчёты точны и соответствуют спецификациям модели (внутренняя валидация модели), что входные и выходные переменные моделей согласуются с имеющимися данными (калибровка модели), и что результаты работы моделей имеют смысл и могут быть объяснены на интуитивном уровне (внешняя валидация модели). В случае, если различные модели одного и того же решения приходят к разным выводам, разработчики модели также должны быть способны объяснить источники различий (кросс-валидация модели). Описание модели должно быть достаточно подробным, так что соответствующие математические расчёты могли бы быть воспроизведены.

Тестирование прогностической валидности, т.е. способности модели делать точные предсказания будущих событий, будет дополнительным плюсом, но не является абсолютно необходимым. Поскольку в будущем может появиться информация, которая не доступна на момент создания и калибровки модели, последняя не должна критиковаться за неспособность предсказать будущее. Тем не менее, хорошая модель должна быть восприимчива к повторной калибровке или уточнению для адаптации к новым данным по мере их поступления. Критерием для определения того, следует ли, и если да, то в какой степени, проводить тестирование модели перед её использованием на предмет достоверности предсказаний, зависит от преимуществ с точки зрения улучшения модели для принятия решений, а также от затрат, связанных с задержкой потока информации при получении дополнительных данных [2].

Оценка качества моделей

Остальная часть данного отчёта описывает согласованное мнение членов Рабочей группы в отношении признаков, которые определяют качественную модель принятия решений в сфере здравоохранения.

Мы активно заимствуем информацию из нескольких превосходных работ, которые предлагают критерии для оценки качества моделей [3—6]. Признаки организованы по основным разделам: структура модели, данные и валидация модели.

Структура модели

1. Модель должна быть построена таким образом, чтобы её входные и выходные параметры имели отношение к позиции, с которой проводится экономическая оценка при принятии решения. Как затраты, так и рассматриваемые медицинские исходы, должны отражать выбранную позицию принятия решений. Например, если исследование призвано помочь лицам, принимающим решения, в распределении ресурсов в широком диапазоне медицинских вмешательств на уровне общества, то результаты модели должны быть широко применимы, и в модель должны быть включены все значимые затраты и последствия для всех членов пострадавшей популяции. Если рассматриваемый ракурс уже, чем социальный, то в отчёте следует отразить, по крайней мере, качественно, последствия расширения модели до социальной точки зрения.
2. Структура модели должна соответствовать как теоретическим положениям, отвечающим за описание моделируемого состояния здоровья, так и имеющимся данным о причинно-следственных связях между переменными. Это не означает, что все причинные связи должны быть доказаны, как это обычно понимается при проверке гипотез, когда показывается, что величина эффекта является статистически значимой при принятом уровне значимости (например, $P < 0,05$). Напротив, это значит, что предполагаемые связи не противоречат имеющимся данным и соответствуют общепризнанным теориям.
3. Если данные в отношении предположений о структуре модели являются неполными, а общепризнанная теория, описывающая процесс протекания рассматриваемого заболевания, отсутствует, то в таком случае следует указать ограничения данных, подтверждающих выбранную структуру модели. По возможности, должен быть выполнен анализ чувствительности с использованием альтернативной структуры модели, например, с использованием альтернативных суррогатных точек или промежуточных переменных.

Пункты 4—8 относятся к моделям переходов между состояниями (камерные модели или модели Маркова):

4. Состояния здоровья могут быть определены так, чтобы они отражали соответствующий процесс протекания заболевания, который может быть

- незаметным или ненаблюдаемым, наблюдаемое состояние здоровья, или же сочетание того и другого. Например, в моделях скрининга состояния здоровья могут определяться, основываясь на основной патологии, клиническом статусе, или на том и другом. Тем не менее, во избежание предвзятости и связанных с этим отклонением в структуре модели, следует соблюдать осторожность, когда вмешательства изменяют как основное заболевание, так и клинические проявления, как, например, в моделях скрининга рака, в которых случаи обнаруженного рака могут иметь разные прогнозы в зависимости от способа или частоты скрининга. В целом, погрешности в структуре модели избегаются путём моделирования состояний, описывающих основное заболевание, с последующей калибровкой выходных значений с учётом данных по наблюдаемым клиническим проявлениям заболевания.
5. В случаях, когда скорости переходов между состояниями или вероятности переходов зависят от событий или состояний, имевших место ранее, такого рода зависимость, или «память», должна быть отражена в модели. Это может быть сделано либо путём включения истории болезни и лечения в определение состояний здоровья, либо путём её включения в качестве ковариаты при определении вероятностей перехода между состояниями.
 6. Состояния не должны быть опущены из-за отсутствия данных. Примерами могут служить хронические состояния здоровья, соответствующие редким побочным явлениям, или осложнения болезни, которые не наблюдаются в клинических исследованиях. Тем не менее, включение в модель определенного состояния здоровья должно основываться на доказательствах в соответствии с вышеупомянутой рекомендацией #2.
 7. Причины включения дополнительного дальнейшего деления состояний здоровья могут быть основаны на их клиническом значении, их отношении к смертности, отношении к качеству жизни или предпочтениях пациентов, их отношении к стоимости ресурсов или любой комбинации вышеперечисленных факторов. Состояния болезни, которые не рассматриваются как клинически значимые, возможно также важно включить в модель отдельно именно по этим причинам. С другой стороны, состояния здоровья, которые рассматриваются как имеющие клиническое значение, могут быть включены для усиления внешней валидности, даже если они не оказывают существенного влияния на результаты модели.
 8. Длина цикла в модели должна быть достаточно короткой, так, чтобы были маловероятны различные изменения в патологии, симптомах, принятии решений о лечении или расходах в пределах одного цикла. Выбор длины цикла должен быть обоснован.
 9. Структура модели, описывающая процесс протекания заболевания и медицинские вмешательства, должна быть как можно более простой. Не обязательно моделировать всю сложность болезни, если решение может быть получено при помощи более простой модели, адекватно описывающей состояния болезни или популяционные подгруппы. При этом, если были сделаны упрощения, их использование должно быть оправдано на основании того, что оно не будет оказывать существенного влияния на результаты анализа. Иногда анализ чувствительности структуры модели, который использует менее обобщенную модель, позволяет удостовериться в том, что упрощения не оказывают существенного влияния на результаты.
 10. Варианты и стратегии не должны быть строго ограничены наличием прямых доказательств из клинических исследований. Также, диапазон моделируемых вариантов не должен ограничиваться лишь теми, которые являются в настоящее время принятой клинической практикой. Должен быть баланс между включением широкого круга возможных вариантов и необходимостью сохранять модель управляемой, интерпретируемой и основанной на фактических данных.
 11. В то время как структура модели должна отражать главные черты заболевания и соответствующие медицинские вмешательства независимо от наличия данных, ожидается, что наличие данных может повлиять на выбор в отношении структуры модели. Например, если конкретная система стадирования используется наиболее часто в клинических исследованиях, то состояния здоровья могут также быть определены в соответствии с этой системой, даже если другие системы стадирования работают лучше с точки зрения прогнозирования результатов или с точки зрения дифференциации качества жизни и затрат.
 12. Отсутствие учёта неоднородности в моделируемой популяции может привести к ошибкам в результатах модели. При необходимости, моделируемые популяции пациентов следует разбивать по группам, которые имеют различные вероятности событий, качество жизни и затраты. Это особенно важно, когда частоты повторяющихся событий с течением времени коррелируют в субпопуляциях, которые имеют различные вероятности наступления событий, так как неспособность сделать это может привести к неправильной оценке долгосрочных результатов.
 13. Временной горизонт модели должен быть достаточно длинным, чтобы отразить важные и значимые различия между долгосрочными последствиями и издержками альтернативных вариантов и

стратегий. Горизонт моделирования, равный по продолжительности времени жизни, подходит для многих моделей и почти всегда требуется для моделей, в которых варианты имеют различные изменяющиеся во времени показатели выживаемости. Использование более коротких горизонтов моделирования может быть обосновано, если выживание и долгосрочные хронические последствия не отличаются среди сравниваемых вариантов или основываясь на понимании процесса протекания болезни и эффективности вмешательств. В любом случае, отсутствие долгосрочных данных наблюдения не должно использоваться как оправдание для отказа расширить временной горизонт до необходимого в рамках рассматриваемого анализа по принятию решения.

Данные

Мы рекомендуем группировать входные параметры моделей по трём категориям: идентификация данных, моделирование данных и применение данных.

Идентификация данных

1. Модель не следует критиковать из-за того, что существующие данные недостаточно отвечают идеальным стандартам научной строгости. Решения будут приниматься в любом случае — с моделями или же без них. Отклонение моделей на основании неполных данных будет означать, что решение, принятое при отсутствии как данных, так и модели, лучше, чем решения с моделью, но без данных. При применении модели имеющиеся данные могут быть использованы для логического обоснования решения. При отсутствии модели будет упущена возможность использовать имеющиеся данные в рамках логической структуры исследования.
2. По ключевым входным параметрам моделей должны проводиться систематические обзоры литературы. Модель должна сопровождаться информацией о том, что такие обзоры были проведены, или же необходимо обосновать отсутствие таких обзоров вследствие адекватности и обобщаемости приобретенных данных.
3. Диапазоны изменения (т.е. верхняя и нижняя границы) должны быть указаны для оценок всех входных параметров базового сценария, для которых выполняется анализ чувствительности. Выбор параметров для проведения анализа чувствительности носит оценочный характер и зависит от мнения аналитика, но неисполнение анализа чувствительности для параметра, значение которого может быть оспорено, может поставить под вопрос все сделанные выводы.
4. В вероятностный анализ чувствительности может быть включена информация по описанию вероятностных распределений для входных параме-

тров на основе неопределенности выборки и/или информации об отклонениях значений в разных исследованиях. Это, однако, не всегда является необходимым или экономически-эффективным. С целью оценки распределений входных параметров предпочтительным вариантом методологии является использование апостериорных распределений, полученных из официальных мета-анализов и байесовского анализа, но практические соображения могут привести и к использованию экспертных оценок (см. ниже, пункт #7).

5. Если известные источники данных исключаются из рассмотрения при оценке параметров, такое исключение должно быть обосновано.
6. Источники данных и результаты не должны быть отклонены только на основании того, что они не достигают общепринятых вероятностных порогов, определяющих «статистическую значимость» (например, $P < 0,05$). Все данные, даже недостаточные для исключения случайности как причины, могут быть на законных основаниях включены в модели. Это действительно при том условии, что неопределенность в отношении оценок указана и проверена в анализе чувствительности, и что в выводах четко указана обусловленность результатов используемыми входными значениями параметров.
7. Экспертное мнение является законным методом оценки параметров при условии, что показано, что эти параметры не оказывают существенного влияния на результаты, либо что по этим параметрам проводится анализ чувствительности с четким указанием, что результаты рассматриваются только в условии этих субъективных оценок. Если результаты чувствительны к используемому мнению экспертов, то процесс сбора информации должен быть детально описан. Предпочтительными являются экспертные оценки, полученные более строгими методами, такими как метод Дельфи или метод номинальных групп.
8. Необходимо убедиться, что перед началом моделирования были рассмотрены разумные возможности для получения новых дополнительных данных. «Разумный» в данном контексте означает, что стоимость и задержки, связанные с получением данных, оправданы ожидаемой ценностью новой информации при анализе. Наряду с существующими формальными методами оценки ценности информации допустимо привести эвристические аргументы относительно того, почему текущая доказательная база была оптимальной с точки зрения обоснования текущих решений. Этого часто можно достичь, применяя анализ чувствительности, чтобы показать, что приемлемые диапазоны данных приведут к качественно схожим выводам, или утверждая, что стоимость и задержки в получении данных не стоят упущенной выгоды от использования только имеющихся данных.

Моделирование данных

1. Моделирование данных представляет собой последовательность математических шагов, направленных на то, чтобы преобразовать эмпирические наблюдения в форму, пригодную для моделей принятия решений. Например:
 - a. Метод для объединения оценок эффективности лечения, взятых из клинических исследований, с базовыми оценками результатов основного сценария, взятыми из эпидемиологических или общедоступных медицинских данных. Оценки эффективности могут быть основаны либо на данных по выборкам пациентов, собирающихся проходить лечение, либо на данных по выборкам пациентов проходящих лечение, в зависимости от целей анализа. Часто уместным подходом является получение оценок относительного риска (или отношения шансов) для различных методов лечения, полученных из клинических исследований, и наложение их на базовые вероятностные оценки (например, в отсутствие лечения или при стандартном лечении) выживания или других ожидаемых результатов из популяционных источников.
 - b. Метод для преобразования интервальных оценок вероятностей из литературных источников или из клинических исследований в моментные показатели, а затем в вероятности перехода или вероятности наступления событий, соответствующие временному интервалу, используемому в модели.
 - c. Метод для объединения в модели показателей смертности по причине болезни и общей смертности. В целом, допустимо получать вероятности общей смертности на основе таблиц продолжительности жизни, используемых в государственной статистике, за исключением случаев, когда может быть обоснован иной источник информации. В общем случае, не требуется делать поправку на то, что смертность от всех причин включает в себя смертность по причине конкретной болезни в общей популяции, если только заболевание не представляет собой основную причину смерти в моделируемых демографических группах.
 - d. Метод моделирования выживаемости (например, в виде экспоненциального, гамма распределения, распространения Вейбулла или Гомперца). Выбор функциональной формы для смертности по причине болезни должен быть указан и обоснован. В общем случае, смертность от всех причин следует моделировать непараметрическими методами на основе данных таблицы смертности.
 - e. Моделирование факторов риска или вмешательства как имеющие аддитивный или мультипликативный эффект на исходные вероятности или частоту заболеваемости или смертности. Доказательства, подтверждающие аддитивную либо мультипликативную форму, следует искать в исследованиях, посвященных изучению воздействию фактора риска или вмешательства на популяцию, стратифицированную по исходному риску.
 - f. Метод для объединения полезностей, специфичных для конкретной предметной области, в многомерную функцию полезности. Лучше при этом использовать проверенные шкалы оценки состояния здоровья и качества жизни с заранее определенными системами оценок, основанными на методах вынужденного выбора (метод стандартного риска, метод временного компромисса).
 - g. Метод для преобразования значений состояния здоровья (например, рейтинговых шкал или классификаций состояний здоровья) в весовые показатели качества жизни.
 - h. Метод для преобразования возможных издержек в затраты.
 - i. Метод корректировки учета инфляции или покупательной способности во времени и между странами. Корректировка с учетом инфляции должна быть основана на индексе потребительских цен (ИПЦ), входящих в него разделах, связанных со здравоохранением, или одним или более подразделах, таких как медицинские услуги или оборудование. Выбор между общим ИПЦ и его компонентами или подкомпонентами, связанными со здравоохранением, зависит от того, в каком из случаев цены на ресурсы представлены лучше: в виде общей «рыночной корзины» в ИПЦ или по потребительской корзине рынка здравоохранения. Ограничением в использовании ИПЦ является то, что он отражает не только цены, но и в некоторой степени величины входных ресурсов, используемых для предоставления медицинских услуг. Предпочтительным методом для внесения корректировок в разных странах является использование паритета покупательной способности. Тем не менее, может быть уместно простое конвертирование валюты, при наличии международного рынка поставок по фиксированной цене.
 - j. Метод дисконтирования затрат и последствий для здоровья.
2. Предположения, положенные в основу модели, должны быть описаны и иметь всеобщее признание, а также, желательно, подтверждение их эмпирической валидности. Ключевые шаги, принятые при разработке модели, должны быть тщательно задокументированы и записаны. Степень доверия к модели может быть повышена, если показано, как модель появилась, например,

- до или во время клинических исследований III или IV фазы, и как ее структура и входные данные видоизменялись с появлением новых данных (например, после завершения клинических исследований) в ответ на последующие обсуждения с клиническими, нормативными и стратегическими экспертами.
3. В случае если альтернативные, но не менее оправданные подходы к моделированию данных могут привести к значительно отличающимся результатам, для оценки роли этих альтернатив необходимо выполнять анализ чувствительности. Например, если модель предсказывает незначительный прирост продолжительности жизни в пожилом возрасте, но при этом модель использует мультипликативный вариант эффекта вмешательства на базовый показатель смертности, то следует также проверить и альтернативу в виде аддитивной модели. Если имеются веские эмпирические доказательства в поддержку одной функциональной формы, то данная форма должна быть использована в базовом сценарии, а альтернативная форма (альтернативные формы) должны быть проверены при проведении анализа чувствительности.
 4. Методы моделирования данных должны соответствовать общепринятым методам, используемым в биостатистике и эпидемиологии. Мета-анализ является допустимым и целесообразным подходом для моделирования, при условии, что приняты меры по выявлению гетерогенности среди источников данных. Гетерогенность можно учитывать либо путем разделения оценок, основанных на различном группировании первичных исследований, либо путем оценки формальных иерархических моделей для объединения информации, полученной из гетерогенных исследований.

Включение данных

1. Единицы измерения, временные интервалы и характеристики исследуемой популяции должны быть взаимно согласованы по всей модели.
2. Допустимы либо вероятностное имитационное моделирование (Монте-Карло, первого порядка), либо детерминированное имитационное моделирование (когортное).
3. В случае использования имитационного моделирования Монте-Карло первого порядка должны быть предоставлены доказательства того, что случайная погрешность моделирования (например, стандартное отклонение для выходных значений за один прогон) заметно меньше, по сравнению с величинами представляющих интерес эффектов.
4. Любое моделирование должно включать в себя подробный анализ чувствительности по ключевым параметрам. Допустимым является проведение детерминированного (одностороннего или многостороннего) либо вероятностного анализа чувствительности.
5. Для минимизации случайной погрешности моделирования, при проведении анализов чувствительности в моделях, которые используют имитационное моделирование по методу Монте-Карло, по возможности следует использовать фиксированное начальное распределение случайных чисел в пределах каждого анализа чувствительности.
6. При использовании когортного моделирования анализ чувствительности может быть произведен при помощи вероятностного имитационного моделирования (Монте-Карло, второго порядка), используя заданные вероятностные распределения входных параметров. При определении этих распределений следует проявлять особое внимание, чтобы гарантировать, что взаимозависимость между параметрами надлежащим образом отражена в совместном распределении параметров.
7. Когда это необходимо, и, если различия в выживаемости с поправкой на качество жизни между альтернативными медицинскими технологиями являются менее длины одного цикла, следует использовать полуциклы для корректировки связанных со временем показателей модели.

Валидация модели

Наши рекомендации по валидации моделей сгруппированы в три категории: внутренняя валидация, межмодельная валидация и внешняя валидация.

Внутренняя валидация

1. Модели должны тщательно тестироваться и проверяться на предмет наличия ошибок и последующего их исправления. Необходимо указывать подтверждения того, что такая внутренняя проверка модели была произведена. Внутренняя проверка должна включать в себя проверку правильности работы модели при задании нулевых или крайних значений входных параметров. Также к такой проверке можно отнести проверку программного кода на наличие синтаксических ошибок, а также тестирование репликации с использованием эквивалентных значений ввода.
2. По возможности, модели должны быть откалиброваны по данным. Калибровка возможна, когда существуют данные как по входным, так и по выходным показателям модели на протяжении всего рассматриваемого промежутка времени. Калибровочные данные могут быть взяты из государственной статистики здравоохранения, такие как общая смертность и смертность по возрастно-половым группам, количество госпитализаций, процедур или стоимость ресурсов. Калибровочные данные должны быть взяты из

источников, не зависящих от данных, используемых для оценки входных параметров модели. Модель не следует критиковать, если не существует независимых калибровочных данных. Тем не менее, модель подвергается критике, если независимые данные, пригодные для валидации, существуют и, либо модель не в состоянии произвести результаты, согласующиеся с этими данными (или несоответствия не могут быть объяснены) либо разработчики модели не исследовали соответствие между результатами моделирования и вышеуказанными независимыми данными.

3. В то время как исходный код должен, как правило, оставаться в собственности разработчика модели, обоснованные просьбы о копиях моделей с соответствующим пользовательским интерфейсом должны быть доступны для экспертной оценки в условиях строгой безопасности и защиты прав собственности.

Межмодельная валидация

1. Модели должны быть разработаны независимо друг от друга для возможности сравнения результатов одной модели с результатами другой модели, моделирующей тот же самый процесс (конвергентная валидность).
2. Если получаемые в модели результаты в значительной мере отличаются от опубликованных или общедоступных результатов, полученных на основе других моделей, разработчик должен предпринять серьезные усилия, чтобы объяснить расхождения. Являются ли расхождения следствием различий в структуре моделей или в значениях входных параметров?
3. Разработчики моделей должны сотрудничать с другими разработчиками при сравнении результатов и формулировании причин возможных расхождений. (Мы приветствуем финансирующие организации, которые поддерживают такой тип сотрудничества, например, программу CISNET моделирования рака при поддержке американского Национального института исследований злокачественных новообразований).

Внешняя и прогностическая валидация

Модели должны быть основаны на лучших данных, имеющихся на момент построения модели. В таких областях, как ВИЧ и гиперлипидемия, более ранние модели предполагали, что последствия для здоровья опосредованы факторами риска (количеством клеток CD4, содержанием сывороточного холестерина). Последующие данные некоторых клинических исследований противоречат оценкам исходных моделей, в то время как другие находятся в соответствии с модельными предположениями.

Информация, полученная из клинических исследований, легла в основу второго поколения моделей, описывающих как ВИЧ, так и гиперлипидемию. Оценки, полученные с помощью этих моделей, более точно соответствуют результатам клинических исследований. В случае с моделями для ВИЧ такой результат был достигнут за счет включения резистентности к антиретровирусным препаратам в оценки действенности лечения, а также включения РНК ВИЧ в качестве маркера вирулентности. В случае с гиперлипидемией это было достигнуто путем моделирования липидных фракций ЛПНП и ЛПВП в качестве факторов риска. Оставшиеся расхождения между прямыми эмпирическими данными и результатами моделирования являются необъясненными. Относятся ли они к артефактам дизайна клинических исследований (например, отбор пациентов, перекрестные стратегии лечения), или к основным биологическим факторам (например, С-реактивный белок и статины, иммунологическое восстановление и антиретровирусная терапия) остается на сегодняшний день неизвестным. Таким образом, модели не только используют известные науке данные на момент разработки модели (в то время, когда там еще могут быть ограниченные данные многолетних наблюдений по новым методам лечения), но они также могут служить основой для сравнения и интерпретации информации из новых исследований. Способность моделей к адаптации к новым данным и новому научному пониманию следует рассматривать не как недостаток, а как сильную сторону модельного подхода.

1. Учитывая то, что модели предназначены для помощи в текущем принятии решений, и результаты моделирования являются действительными только в условиях заданных предположений, нет необходимости в проверке каждой оценки и предположения о структуре модели на перспективных исследованиях перед ее использованием.
2. Решение для получения дополнительных данных для добавления в модель должно быть основано на соотношении между ожидаемой ценностью дополнительной информации и стоимостью информации.
 - а. «Ожидаемая ценность информации» относится к одному из основных понятий теории принятия решений, согласно которому информация оценивается с точки зрения ее ожидаемого (или среднего) влияния на последствия решений. Например, ожидаемая ценность информации была бы равна нулю для изучения параметра модели, априорный диапазон значений которого не включает в себя порог для выбора среди вариантов решения. Оценка, связанная с априорными вероятностями возможных результатов исследования является неизбежной частью оценки «ожидаемой ценности информации».

- b. «Стоимость информации» включает в себя стоимость ресурсов для выполнения эмпирического исследования, а также размер ожидаемой упущенной выгоды от затягивания момента принятия решения до завершения исследования. Оценка, связанная с априорными вероятностями эффектов от лечения, является неизбежной частью оценки стоимости информации.
- c. Рекомендации по проведению или дизайну исследований для формирования руководящих принципов в принятии решений могут быть основаны на формальном анализе ценности информации или на неформальной интерпретации выводов анализа чувствительности.
3. Модели никогда не должны считаться полными или неизменными. Они должны систематически обновляться. Иногда от существующих моделей следует отказываться и заменять другими по мере поступления новой информации, которая может быть использована при построении модели или при задании входных значений. Также, следует прекратить пользоваться моделями, которые оказались несовместимыми с обновленными данными и не были скорректированы с учётом последних, до тех пор, пока такая коррекция не будет произведена.

Заключительные комментарии

Хотя данные рекомендации и отражают точку зрения Рабочей группы на момент составления отчёта, они не должны рассматриваться как окончательные и неизменные. Они не являются итоговым «сводом правил». Различные обстоятельства могут привести к отклонениям от описанных рекомендаций, в зависимости от имеющихся в распоряжении разработчика модели ресурсов (времени, денег и данных) и от цели моделирования.

На наш взгляд, самое главное, что следует иметь в виду при анализе модели оценки технологий здравоохранения, это тот факт, что результаты моделирования не должны рассматриваться как утверждения о фактах или как прогнозы будущего. Скорее, цель моделирования заключается в синтезе доказательств и предположений таким образом, чтобы позволить

конечным пользователям получить представление о роли этих входных данных для оценивания последствий и затрат. Результаты моделирования всегда зависят от входных данных, поэтому так важно, чтобы данные на входе модели были настолько прозрачными и доступными, насколько это возможно.

Дальнейшее чтение по методологии моделирования

Цель данного отчёта заключается не в предоставлении некоего обзора методологии моделирования, а скорее в определении тех аспектов методологии, которые Рабочая Группа относит к надлежащей практике проведения исследований. Для читателей, желающих ознакомиться с основами методов моделирования, мы рекомендуем следующие источники. В качестве вводного учебного пособия по теории принятия решений, в том числе деревьям решений и моделям Маркова, смотрите *Hunink* и соавт. [7]. Для ознакомления с современными методами моделирования при проведении экономической экспертизы, включая обзор методов моделирования выживаемости по данным клинических исследований, а также обзор детерминированных и стохастических подходов к моделированию, смотрите Kuntz и Weinstein [8]. Обзор методов обработки неопределенности в моделях представлен в работе *Briggs* [9] и одиннадцатой главе книги *Hunink* и соавт. [7].

Благодарности

Рабочая группа выражает благодарность следующим членам ISPOR за предоставленные ценные письменные замечания по проектам данного отчёта: Phantipa Sakthong, MS, факультет фармацевтических наук, Университет имени Короля Чулалонгкорна, Бангкок, Таиланд; Mendel Singer, PhD, университет Кейс Вестерн Резерв, Кливленд, Огайо, США, и Leslie Wilson, PhD, MS, Калифорнийский университет в Сан-Франциско, Калифорния, США.

Авторы также благодарят Исполнительного Директора ISPOR Marilyn Dix Smith за оказанную организационную поддержку в созыве совещаний Рабочей Группы.

Литература

1. National Research Council. Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling. Vol. 1, Review and Recommendations. Washington: National Academy Press; 1991.
2. Weinstein MC, Toy EL, Sandberg EA, et al. Modeling for health care and other policy decisions: uses, roles, and validity. *Value Health* 2001; 4: 348—61.
3. Sculpher M., Fenwick E., Claxton K. Assessing quality in decision analytic cost-effectiveness models: a suggested framework and example of application. *Pharmacoeconomics* 2000; 17:461—77.
4. Hay J., Jackson J., Luce B., et al. Methodological issues in conducting pharmacoeconomic evaluations—modeling studies. *Value Health* 1999; 2:78—81.
5. Akehurst R., Anderson P., Brazier J., et al. Decision analytic modeling in the economic evaluation of health technologies. *Pharmacoeconomics* 2000;17: 443—4.
6. Gold MR, Siegel JE, Russell LB, Weinstein MC, editors. Cost-Effectiveness in Health and Medicine. Report of the Panel on Cost-Effectiveness in Health and Medicine. New York: Oxford University Press; 1996.
7. Hunink M., Glasziou P., Siegel J., et al. Decision Making in Health and Medicine: Integrating Evidence and Values. Cambridge: Cambridge University Press; 2001.
8. Kuntz K., Weinstein M. Modelling in economic evaluation. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001.
9. Briggs A. Handling uncertainty in economic evaluation and presenting the results. In: Drummond M, McGuire A, editors. Economic Evaluation in Health Care: Merging Theory with Practice. Oxford: Oxford University Press; 2001.

Многонациональные исследования — Рекомендации по требованиям к переводам, подходам к использованию одного языка в разных странах, а также к подходам в поддержке при сборе данных: отчет Целевой группы ISPOR по надлежащей исследовательской практике в переводах и лингвистической валидации оценок результатов пациентами

Diane Wild, MSc¹, Sonya Eremenco, MA², Isabelle Mear, MA³, Mona Martin, RN, MPA⁴,
Caroline Houchin, MA¹, Mary Gawlicki, MBA⁵, Asha Hareendran, PhD⁶, Ingela Wiklund,
PhD⁷, Lee Yee Chong, PhD⁸, Robyn von Maltzahn, MSc¹, Lawrence Cohen, PharmD, BCPP,
FASHP, FCCP⁹, Elizabeth Molsen, RN¹⁰

¹ — Oxford Outcomes Ltd, Oxford, UK

² — United BioSource Corporation, Bethesda, MD, USA

³ — Mapi Research Institute, Lyon, France

⁴ — Health Research Associates, Inc., Mountlake Terrace, WA, USA

⁵ — Corporate Translations, Inc., Hartford, CT, USA

⁶ — Global Outcomes Research, Pfizer, Ltd., Sandwich, UK

⁷ — Global Health Outcomes, GlaxoSmithKline, Middlesex, UK

⁸ — Formerly at AHP Research Ltd, Middlesex, UK

⁹ — Washington State University, Spokane, WA, USA

¹⁰ — International Society for Pharmacoeconomics & Outcomes Research (ISPOR), Lawrenceville, NJ, USA

[Correction added after online publication 13-November-2008: Information on FDA requirements in Section I has been updated]

Переводчик:

Павлыш Андрей Владиславович — к.м.н., докторант Первого Санкт-Петербургского государственного медицинского университета им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Вербицкая Елена Владимировна — к.б.н., доцент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Резюме. *Цели:* интернационализация программ клинических исследований повысила необходимость переводить и культурно адаптировать оценки результатов пациентами (PRO). Хотя существуют руководства по надлежащей практике перевода и лингвистической валидации, Целевая группа ISPOR по переводам и лингвистической валидации при оценке результатов пациентами определила ряд областей, в которых, по мнению ее участников, требовалось дальнейшее обсуждение методов и надлежащих практик. Группа выделила следующие области: 1) выбор языков, необходимых для многонациональных исследований; 2) подходы, предлагаемые, если в двух или более странах используется один и тот же язык; и 3) оценка эквивалентности измерений для объединения данных из разных стран. *Методы:* Целевая группа рассмотрела эти три области, прочитала имеющуюся литературу и провела множество обсуждений для разработки данного отчета. *Результаты:* Инструменты для помощи в принятии решений также были разработаны и представлены для выбора языков и подходов, предлагаемых для использования одного и того же языка в разных странах. *Заключение:* Мы надеемся, что этот отчет и предлагаемые инструменты для принятия решений помогут участникам многонациональных исследований: 1) принимать решения о переводах, необходимых для конкретной страны; 2) выбирать подход, когда на одном и том же языке говорят больше, чем в одной стране; и 3) выбирать методы для сбора данных, полученных с помощью различных языковых версий одного и того же инструмента.

Ключевые слова: адаптация, лингвистическая валидация, многонациональный, объединение, перевод

Multinational Trials — Recommendations on the Translations Required, Approaches to Using the Same Language in Different Countries, and the Approaches to Support Pooling the Data: The ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force Report

Abstract. *Objectives:* With the internationalization of clinical trial programs, there is an increased need to translate and culturally adapt patient-reported outcome (PRO) measures. Although guidelines for good practices in translation and linguistic validation are available, the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Task Force identified a number of areas where they felt that further discussion around methods and best practices would be beneficial. The areas identified by the team were as follows: 1) the selection of the languages required for multinational trials; 2) the approaches suggested when the same language is required across two or more countries; and 3) the assessment of measurement equivalence to support the aggregation of data from different countries. *Methods:* The task force addressed these three areas, reviewed the available literature, and had multiple discussions to develop this report. *Results:* Decision aid tools have also been developed and presented for the selection of languages and the approaches suggested for the use of the same language in different countries. *Conclusion:* It is hoped that this report and the decision tools proposed will assist those involved with multinational trials to 1) decide on the translations required for each country; 2) choose the approach to use when the same language is spoken in more than one country; and 3) choose methods to gather evidence to support the pooling of data collected using different language versions of the same tool.

Keywords: adaptation, linguistic validation, multinational, pooling, translation

Введение

Совет ISPOR по здоровью, науке и политике одобрил и рекомендовал создание Целевой группы по переводу и лингвистической валидации в январе 2007 года. Совет ISPOR одобрил это в марте 2007 года.

Члены Целевой группы являются опытными и хорошо осведомленными специалистами в области перевода, лингвистической валидации и измерению международной эквивалентности, и работают в научной среде, промышленности, контрактных исследовательских организациях (contract research organizations, CRO), а также в качестве советников в правительстве. Они представляют несколько стран Европы, а также США.

В рамках форума по оценкам результатов пациентами (patient-reported outcome, PRO), проводившегося на 12-й ежегодной международной встрече ISPOR в 2007 году, был представлен обзор первоначальных рекомендаций и будущего направления работы Целевой группы. На форуме была также получена обратная связь.

Целевая группа встречалась раз в месяц, чтобы обсудить наиболее важные вопросы, возникшие в связи с переводами и лингвистической валидацией документов PRO. Целевая группа была разделена на три подгруппы, которые разрабатывали наброски и проекты отчетов, каждая по своей соответствующей теме. Все результаты работы были рассмотрены Целевой группой в полном составе.

После того, как был готов проект версии окончательного отчета, состоящего из трех разделов, он был представлен на рассмотрение в течение трех недель группе особого интереса по оценкам результатов пациентами. Была получена важная и конструктивная обратная связь, которую, по мере необходимости, включили в отчет. Кроме того, содержание отчета было представлено в 2008 году на 13-й ежегодной международной конференции ISPOR о PRO. Вновь полученные комментарии включались в окончательный отчет. Члены Целевой группы также получили многочисленные варианты отчета за время его 18-месячной разработки. После того, как было достигнуто соглашение, рукопись предоставили журналу Value in Health.

Растущая интернационализация клинических исследований вызвала необходимость переводов и адаптации инструментов PRO для использования в странах, чей язык отличается от языка оригинала, и она продолжает расти с привлечением в клинические исследования новых стран, таких как Индия и Китай. Большинство инструментов разрабатывается в англоговорящих странах, поэтому они должны быть переведены и адаптированы для использования в других странах.

Эта Целевая группа была основана в марте 2007 года с целью распространения принципов надлежащей практики по переводу и процессу культурной адаптации оценок результатов пациентами (PRO): Отчета Целевой группы ISPOR по переводу и культурной адаптации [1]. Члены Целевой группы представляют научные круги, службу лингвистической валидации и промышленность.

Целевая группа вычленила три важных вопроса, которые не были рассмотрены в предыдущих отчетах или в литературе: 1) выбор языков, необходимых для перевода; 2) методы перевода для одноязычных версий, используемых в разных странах; и 3) вопросы и проблемы объединения данных PRO для разных стран.

РАЗДЕЛ 1 — Многонациональные исследования: помощь в принятии решений о выборе языков для перевода

Введение

Поскольку клинические исследования проводят во всё большем числе стран, становится необходимым выбрать языковые версии, требующиеся для эффективного использования PRO. Нужно найти компромисс между необходимостью адекватно покрыть все используемые целевой популяцией языки, быть эффективными для целевой группы населения и необходимостью экономить ресурсы.

Важно не недооценить требуемое время, ресурсы и сложность реализации внедрения конечных точек PRO среди целого ряда культурно разных стран. Для индустрии крайне важно адекватно оценить значимость этих вопросов (выбор страны, селекционный центр, число пациентов на сайте, необходимость и стоимость обучения и т.д.). Чем больше пациентов можно набрать через один сайт, тем ниже риск ошибок в данных из-за различий при включении сайтов в исследование. Это особенно важно для развивающихся стран, где может не быть стандартизированных критериев в здравоохранении или процедур. Меньшее число клинических сайтов с большим количеством пациентов обеспечит большую согласованность процедур исследования.

Кроме того, может оказаться разумным рассмотрение набора относительно большого количества пациентов в пределах данной страны (например, 100 пациентов). Во-первых, это сделает психометри-

ческую оценку инструментов более осуществимой. Во-вторых, могут существовать нормативные правила, специфические для данной страны, по которым в клиническом исследовании должно принять участие определенное количество пациентов внутри страны, чтобы продукт был одобрен для использования. Регуляторные вопросы также должны быть приняты во внимание при определении того, сколько пациентов набирать в каждой стране.

Целью этого раздела отчета было определение процесса выбора количества языков, необходимых для адекватной работы с глобальными группами людей для исследований.

Методы

Подгруппа раздела 1 сосредоточила свои усилия на разработке инструмента, который будет оказывать помощь в определении того, какие языки необходимы для стран, отобранных для участия в глобальном клиническом исследовании. Цель состояла в том, чтобы собрать необходимую информацию и предложить процесс принятия решений, которые будут направлять пользователей к оптимальному варианту, а не к окончательным решениям.

Поиск литературы мало что выявил по этому вопросу. После его завершения, в качестве первого проекта инструмента была разработана Сетка (форма) высокой производительности, для чего потребовалась следующая информация:

1. Анализ группы населения: определить происхождение популяции и ее потенциальное влияние на языки, используемые в этой стране.
2. Распространенность заболевания в стране: является ли это фактором, влияющим на выбор языка.
3. Анализ необходимости включения языка: предоставить одновременно информацию об исследовании, которая может повлиять на выбор языка, и информацию о языке, которая должна быть принята во внимание при принятии решения о включении.

Для того чтобы оценить простоту использования и согласованности с точки зрения результатов и ясности инструкций, три компании лингвистической валидации использовали сетки (таблицы-формы), чтобы выбрать язык для следующих стран: Аргентина, Чили, Южная Африка, США, Индия и Сингапур. Результаты согласовывались между тремя компаниями.

Разбирательство после использования сетки (формы) показало, что инструкции были ясны, но потребовались некоторые изменения для подтверждения и уточнения необходимой информации. Кроме того, было разработано дерево решений (рис. 1), чтобы продемонстрировать, каким образом следует проводить процесс выбора после того, как была заполнена Сетка (форма-таблица). См. финальную форму (табл. 1).

Сетка (таблица-форма) выбора языка

Источники информации:					
Цель:	определить происхождение населения и его влияние на языки, на которых говорят в стране				
1. Анализ населения					
Цель: определить происхождение населения и его влияние на языки, на которых говорят в стране					
% населения, рожденного в стране	% иммигрантов, не рожденных в стране	Происхождение иммигрантов	Уровень грамотности	Национальный официальный язык(и)	Язык правительства и СМИ
×	×	×	×	×	×
2. Преобладание заболевания					
Этничность	Возрастные рамки	Социоэкономическая группа			
×	×	×			
3. Анализ необходимости включения языка					
Цель: определить вероятность включения языка согласно критериям, указанным ниже.					
Страна:	XXXX				
Заполняется персоналом			Город или регион	Город или регион	Город или регион
Определенное количество пациентов на сайт					
Заполняется персоналом			Спикеры языка А	Спикеры языка В	Спикеры языка С
			×	×	×
Расположение сайтов исследования					
Ассоциированный преобладающий язык(и)					
Главный язык(и), определенный кол-вом спикеров			Язык	Язык	Язык
Официальный язык (Да/Нет)			×	×	×
Устный/письменный язык			×	×	×
% и количество людей, говорящих на этом языке как на родном			×	×	×
Изначальное расположение людей, говорящих на этом языке (укажите любое расположение)			×	×	×
Вывод: нужен язык или нет			×	×	×

Сетка (форма) должна быть кем-то заполнена данными о языках, на которых говорят в целевой стране, а для заполнения части 3, раздел 1 (анализ необходимости включения языка) необходимо знание дизайна исследования. Самой большой проблемой, встречающейся при выборе языка, является точность и доступность данных. Вся информация должна быть получена из достоверных источников (если это возможно, то из государственных регулярно обновляемых сайтов). Любая информация, включенная в форму, должна иметь перекрестные ссылки и быть явно указана в начале документа. В идеале, нужно использовать три разных источника, и далее либо сохраняется самый надежный показатель, либо рассчитывается средний показатель, если отсутствуют государственные сайты.

Результаты

Решения, касающиеся выбора языка, являются результатом целого комплекса факторов, например, является ли язык официальным (т.е. имеет ли язык юридический статус). Как правило, официальный язык используется в правительстве, в судах и для административных вопросов. В большинстве случаев выбирается официальный язык, если число носителей не ограничено. Тем не менее, порог может быть выражен в процентах по отношению ко всему населению страны. Так, 3 % — ный порог, показанный на Рисунке 1, не представляет собой абсолютный уровень; вместо этого он обеспечивает индикацию уровня, ниже которого включение языка не имеет смысла (согласно предварительному тестированию). Этот уровень может быть ниже или выше в зависимости от

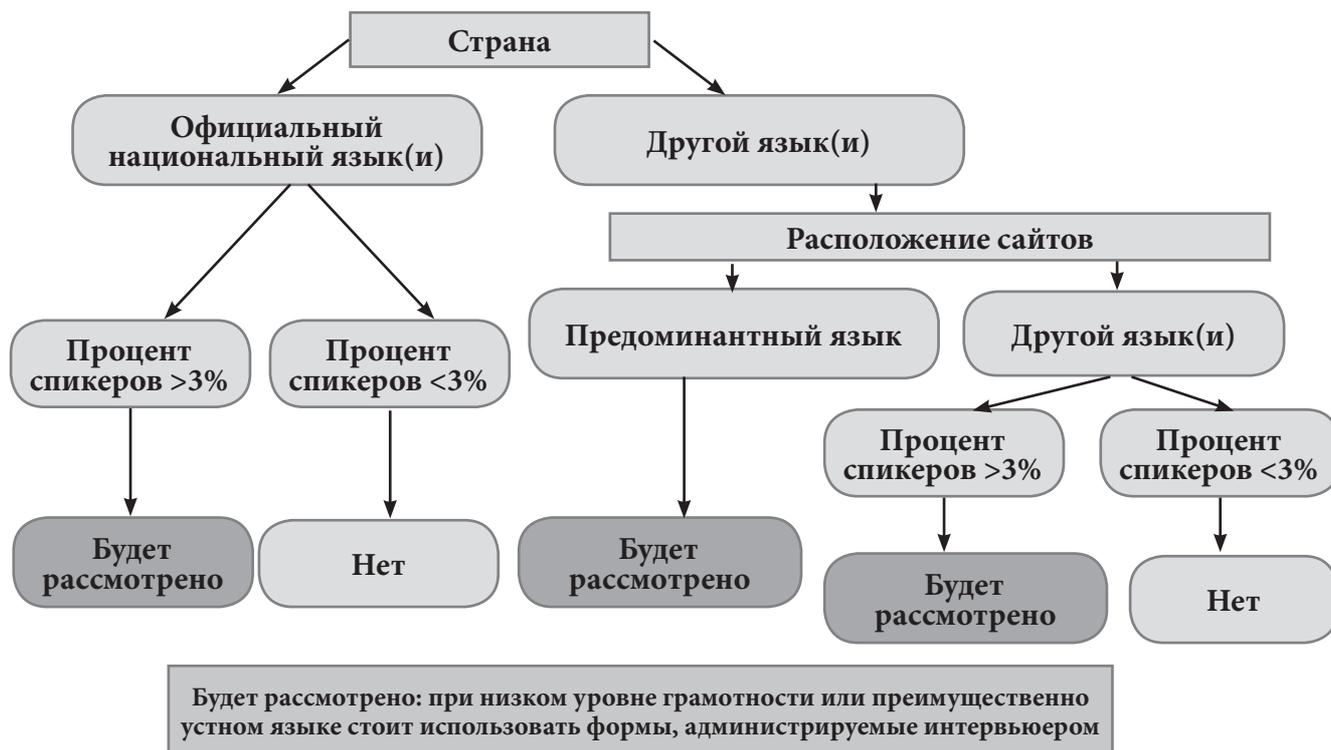


Рис. 1. Древо решений

количества носителей языка, которое он представляет, распространенность заболевания среди населения, и легкость или сложность нахождения пациентов, отражающих заранее определенное целевое население для исследования. Авторы не рекомендуют создавать переводы на языки, на которых говорит всего лишь 3% населения, за вычетом отдельных редких случаев.

Выбор языка для страны

Заключение

Определение того, какие языковые версии необходимы для каждой страны-участницы в клиническом исследовании, остается сложным вопросом, по которому не существует никакого волшебного рецепта. Использование одних и тех же инструментов всеми, кто принимает решения, может значительно облегчить процесс принятия решений и разработать более глубокое обоснование для этой цели.

РАЗДЕЛ 2 — Один язык, разные страны

Введение

После определения, какие языки будут необходимы для каждой страны-участницы в клиническом исследовании, возникает вопрос о том, какой подход является наиболее подходящим для создания необходимых переводов. Общие руководства по переводу инструментов PRO, в целом, были предложены

[1, 2]. Тем не менее, эти руководства фокусируются на общем процессе разработки лингвистически утвержденных переводов, вне зависимости от того, о каком языке идет речь, и не рассматривается вопрос о том, как подойти к проблеме создания инструмента на том же языке для использования в различных целевых странах. Целью данного раздела является определение возможных подходов к обработке этой ситуации и определение соответствующих преимуществ и недостатков каждого подхода, чтобы можно было принимать более обоснованные решения. Решение региональных различий, которые могут возникнуть внутри страны, выходит за рамки этой статьи. Для целей этой статьи мы будем использовать термин «страна» для обозначения различных мест, где используется один и тот же язык.

Справочная информация

Две основные причины, по которым на одном и том же языке могут говорить в разных странах — колониализм и изменения границ (данная фраза не корректна для России). Совсем недавно возросшая иммиграция привела к ситуации, когда в стране с одним изначальным языком все еще в такой мере распространен язык иммигрантов, что услуги для иммигрантов должны быть предоставлены на их языке (например, для турецкого населения в Австрии). В результате географической дисперсии различных языковых групп разговорный язык развился до такой степени, что произношение и словарный запас от-

личаются, (например, португальский в Португалии против португальского в Бразилии). Письменность, как правило, развивается медленнее. Различные языковые группы населения скорее поймут один и тот же письменный язык в отличие от разговорного, что делает возможным использование в различных странах опросников, написанных на том же языке.

Методы

Был проведен обзор литературы с целью изучения существующих правил и подходов, применяемых для решения вопроса переводов, требующихся при использовании одного языка в различных странах. Несколько публикаций отвечает на этот вопрос [2, 3]. Во многих случаях, в публикациях представлены рекомендации, предназначенные для конкретного инструмента, например, FACIT [4, 5], EORTC [6], EQ-5D (руководства EuroQol), и [7], Ноттингемский профиль здравоохранения, но там не было рекомендаций для инструментов PRO в общем. Результаты этого обзора литературы подтвердили необходимость получения дополнительной информации в этой области.

Возможные подходы

Основываясь на литературе и нашем практическом опыте, мы определили три основных подхода к решению проблемы языков в более чем одной стране. Четвертый подход, подразумевающий разработку перевода для использования в одной стране и предположение, что он будет приемлем для использования во всех других странах, разделяющих один и тот же язык, без дальнейшей работы не будет обсуждаться в этой статье. Мы считаем такой подход нецелесообразным, за исключением редких случаев, когда размер выборки по стране настолько мал, что дополнительная работа не является экономически целесообразной. Для целей настоящей статьи будем считать, что английский язык является основным языком инструментов PRO, которые будут переведены. Существуют следующие подходы:

1. Подход, специфический для страны: разные версии перевода разрабатываются для каждой крупной страны или субпопуляции внутри страны.
2. Подход адаптации того же языка: языковая версия документа существует для одной страны; она адаптирована для использования в новых странах или группах населения.
3. Универсальный подход: перевод с самого начала предназначен для нескольких языковых стандартов. Переводчики из разных стран происхождения достигают компромисса для получения перевода, понятного всем.

В некоторых случаях может быть принят комбинированный подход, в котором перевод инициируется с помощью универсального подхода, и, если разрешение разногласий невозможно, реализуются небольшие вариации для каждой страны. Кроме того, можно обратить вспять процесс от специфического для страны к универсальному. Тем не менее, это может оказаться более сложным, потому что перевод изначально ориентирован только на одну группу. Мы обнаружили, что на практике могут быть приняты многие вариации каждого из этих подходов в зависимости от рассматриваемого документа и истории его перевода. Мы разработали ряд сценариев, чтобы лучше проиллюстрировать, как эти подходы могут быть применены в реальных ситуациях. Эти сценарии представляют подходы на всей их протяженности от полностью специфического для страны на одном конце, до полностью универсального на другом конце, с несколькими вариациями подходов, в том числе адаптацию одного языка, между двумя крайностями.

Эти сценарии также включают количество вовлеченных переводчиков, и относительного ранжирования времени и связанных с этим расходов, (основываясь, прежде всего, на количестве переводчиков и типе, и количестве этапов), так же, как и другие способы сравнения различных подходов. В таблице 2 представлены сценарий 1, где есть несколько испаноговорящих целевых стран, и нет существующих доступных переводов. В таблице 3 представлен сценарий 2, где существует перевод для Франции и требуется несколько переводов для франкоговорящих целевых стран. Дополнительные сценарии находятся за рамками данной статьи. Они заключаются в следующем:

1. Через один год после того, как проверка, описанная в сценарии 1, была завершена, потребовалась версия для перуанского испанского.
2. Испанский вариант для Мексики, США, Гватемалы, Колумбии, Венесуэлы и Аргентина является обязательным. В настоящее время испанских переводов не существует.
3. Английский вариант для Канады, Великобритании, Австралии, Индии и Южной Африки является обязательным. Оригинальная версия инструмента написана на американском английском.
4. Китайский вариант для Тайвани и Гонконга является обязательным. В настоящее время переводов не существует.

Примечание: Ограничение места не позволяет воспроизвести в этой статье все сценарии. Если вы хотели бы получить полную версию сценарных таблиц, пожалуйста, свяжитесь с авторами этого раздела.

Таблица 2

Сценарий 1. Испанский для разных испаноговорящих стран: Испания, США, Мексика и Аргентина.
Сейчас не существует перевода на испанский

Номер опции	Описание опции	Плюсы	Минусы	Примерная стоимость	Примерное затраченное время
1. Специфический для страны	<p>Создает четыре разных перевода вне зависимости друг от друга. Разбирается в каждой стране.</p> <p>Требуемые ресурсы: 8–16 переводчиков / 2–4 запасных переводчика</p> <p>Результат: 4 перевода для каждой конкретной страны</p>	<p>1) Все переводы подгоняются под каждую конкретную страну</p> <p>2) Технологические термины, системы образования/здравоохранения, демографические вопросы, названия продуктов и т.п. не потребуют дальнейшей адаптации.</p> <p>3) Переводчики из разных стран не должны будут искать компромиссы.</p> <p>4) Переводчики из разных стран не будут иметь затруднений с разными словарями и структурами предложений в процессе перевода для другой страны.</p>	<p>1) Максимальный потенциал для стилистических и прочих различий в финальном переводе.</p> <p>2) Требуется пользователь для поддержки максимального количества версий.</p> <p>3) Потенциал возникновения путаницы из-за по-разному сформулированных вопросов.</p>	5	1
2. Комбинация специфического для страны и адаптации того же языка	<p>Создает перевод для Испании и разбирается в Испании.</p> <p>Независимо создает второе перевод для любой другой и остальных трех стран. Потом переводчиков из оставшихся двух стран просят адаптировать изначальный перевод для их целевой страны. Разбирательство происходит отдельно в каждой из этих трех стран.</p> <p>Требуемые ресурсы: 6–8 переводчиков / 4 запасных переводчика</p> <p>Результат: 2 перевода для каждой страны / 2 адаптации с одного языка</p>	<p>1) Перевод для Испании и второй перевод подгоняются под конкретную страну.</p> <p>2) Для остальных стран делается лингвистический и культурно адекватная адаптация.</p> <p>3) Возможность адаптировать технические термины, системы образования/здравоохранения, демографические вопросы, названия продуктов и т.п.</p> <p>4) Меньше стилистических и прочих незначимых различий между разными версиями, потому что их независимо адаптируют.</p> <p>5) Освещает проблемы, связанные с адаптацией других версий из версии с известными расхождениями в структуре предложения и использовании.</p>	<p>То же, что и пункты 2 и 3 в предыдущем, плюс:</p> <p>1) Переводчики, делающие адаптации, могут ошибиться из-за оригинальной версии перевода и/или не внести необходимые культурные различия. В этом случае некоторая терминология в получившемся переводе будет некорректной.</p>	4	2

Таблица 2 (продолжение)

<p>3. Комбинация специфического для страны и адаптации того же языка</p>	<p>Создает перевод для Испании. Предоставляет переводчикам из других трех целевых стран испано-испанскую адаптацию и просят их адаптировать перевод для каждой из своих стран. Разбирательство происходит отдельно во всех четырех странах. Требуемые ресурсы: 5–10 переводчиков / 2–4 запасных переводчика Результат: 1 перевод для страны/3 адаптации с одного и того же языка</p>	<p>Те же плюсы, что и пункты (2–5) выше, плюс: 1) Перевод для Испании подгоняется конкретно под эту страну</p>	<p>То же, что пункты 2 и 3 из первого и пункт 1 из второго, плюс: 1) Сложнее делать адаптации из версии на испанском из-за разницы в словарях и построении предложений.</p>	3	2
<p>4. Комбинация специфического для страны и универсального</p>	<p>Создает перевод для Испании и разбирается в Испании. Переводчики из Мексики, США и Аргентины работают вместе, чтобы создать единый перевод, который они посчитают приемлемым в трех целевых странах. Разбирательство происходит отдельно в каждой из стран. Требуемые ресурсы: 5–8 переводчиков / 2–4 запасных переводчика Результат: 1 перевод для страны / 1 универсальный перевод</p>	<p>Тот же плюс, что и 3–1, плюс: 1) Нужно меньше пользователей для поддержки версий 2) Стилистических и прочих расхождений тоже меньше, потому что вторую версию разрабатывают универсально</p>	<p>То же, что пункты 1–3 и 2–1, плюс: 1) Скорее всего, не вся лексика будет одинаково уместной и подходящей для всех стран. 2) Меньше возможностей адаптировать технические термины, системы образования/здравоохранения, демографические вопросы, названия продуктов и т.п. 3) Для достижения соглашения по финальному переводу переводчикам придется идти на компромисс 4) В зависимости от результатов разбирательства все равно могут понадобиться 4 разных версии.</p>	2	1
<p>5. Универсальный</p>	<p>Переводчики из Испании, Мексики, США и Аргентины работают вместе для создания одного перевода, который сочтут приемлемым в четырех целевых странах. Разбирательство происходит отдельно в каждой из четырех стран. Требуемые ресурсы: 4 переводчика / 1–2 запасных переводчика Результат: 1 универсальный перевод</p>	<p>1) Нужны пользователи для поддержки только одной версии 2) Одна версия освещает проблемы стилистических или прочих неважных различий между версиями 3) Меньше возможностей для ошибки, потому что всем задается один и тот же вопрос</p>	<p>То же, что в пунктах 3–3 и 3–4, плюс: 1) Скорее всего, не вся лексика будет одинаково уместной и подходящей для всех стран. 2) Нет возможности адаптировать технические термины и т.д. 3) Сложнее достичь соглашения по лексике 4) Решения о том, подходит ли лексика, принимаются единолично переводчиком в каждой стране</p>	1	2

Таблица 3

Сценарий 2. Много целевых стран: Франция, Бельгия, Канада и Швейцария.
Для Франции существует полностью согласованный и разобраный перевод инструмента А

Номер опции	Описание опции	Плюсы	Минусы	Примерная стоимость	Примерное затраченное время
1. Специфический для страны	Создает три разных перевода вне зависимости друг от друга и существующего перевода для Франции. Разбирается в каждой стране. Требуемые ресурсы: 6–12 переводчиков / 3–6 запасных переводчика Результат: 4 перевода для каждой конкретной страны	Те же, что в сценарии 1, опции 1	Те же, что в сценарии 1, опции 1	4	1
2. Некоторая языковая адаптация	Предоставляет переводчикам из других трех стран франко-французскую адаптацию перевода и просит их адаптировать этот перевод для каждой из их целевых стран самостоятельно. Требуемые ресурсы: 3–6 переводчиков / 40–3 запасных переводчика Результат: 1 перевод для каждой страны/3 адаптации с одного языка	Те же, что 2–2, 2–3, 2–4	Те же, что 1–2, 1–3 и 2–1, плюс: 1) Сложнее создать адаптацию для Канады из-за различий в использовании отдельных слов.	3	4
3. Адаптация языка к региону	Страны в группе в соответствии с «лингвистической схожестью» разрабатывают региональные переводы – один для Европы, один для Канады. Переводчиков из Бельгии и Швейцарии просят адаптировать перевод для использования в каждой из стран. То же для Канады. Разбирательство проходит в Европе большим размером выборки, чтобы обе целевые страны были полностью задействованы. В Канаде разбирательство как обычно. Требуемые ресурсы: 3–6 переводчиков / 0–2 запасных переводчика Результат: 1 перевод для страны / 2 адаптации (движение в сторону универсального)	Те же плюсы, что 2–4 и 3–1	Те же, что в 2–3, 3–1, 3–2, 3–3, 3–4	2	2
4. Универсальный	Переводчики из Франции, Бельгии, Канады и Швейцарии работают вместе для создания одного перевода, который сочтут приемлемым в четырех целевых странах. Разбирательство происходит отдельно в каждой из четырех стран (включая Францию, если получившийся перевод отличается от оригинальной французской версии) Требуемые ресурсы: 4–8 переводчиков / 1–2 запасных переводчика Результат: 1 универсальный перевод	Те же, что в сценарии 1, опции 5	То же, что в сценарии 1, опции 5	1	3

Резюме преимуществ и недостатков подхода, специфического для страны, выглядит следующим образом

Преимущества:

- Позволяет получить дополнительное разговорное и идиоматическое применение. Существенно снизится риск того, что пациенты неправильно поймут используемую формулировку, в том числе термины, используемых для культурных ссылок, таких как образование. Это может быть особенно актуально для пожилых пациентов, которые привыкли к более традиционному языку с меньшим количеством слов, заимствованных из других стран.
- Сроки могут быть короче, чем с универсальным подходом, потому что время не тратится на ожидание соглашения от всех представителей стран.

Недостатки:

- Основные изменения в переводах одного и того же языка могут привести к систематической ошибке и отразить стилистические различия переводчиков, а не истинные различия между языками, которые на самом деле существуют в этих странах. Трудно представить доказательства того, что является одним из важнейших изменений, основанным на реальных лингвистических культурных вопросах, а что — личным мнением участвующих переводчиков.
- Расходы спонсора на печать Индивидуальных Регистрационных Карт, ИРК (case report forms, CRF), могут быть выше.

Резюме преимуществ и недостатков подхода адаптации того же языка выглядит следующим образом

Преимущества:

- При использовании для конкретной страны преимущества аналогичны преимуществам в конкретных странах выше. Он также допускает терминологию, определенную в этой стране, при сохранении части предыдущего перевода.

Недостатки:

- В случаях, когда вовлечено мало переводчиков, может быть увеличено влияние личного мнения, превосходящего лингвистическую необходимость.

Резюме преимуществ и недостатков универсального подхода выглядит следующим образом

Преимущества:

- Всем группам задают один и тот же вопрос; следовательно, меньше вероятность появления по-разному сформулированных вопросов.

- Могут быть практические и логистические преимущества при единственной доступной языковой версии.

Недостатки:

- Формулировка может звучать для пациентов неестественно, и язык может быть в культурном и лингвистическом смысле сглаженным (однако, в большинстве случаев, пациенты в состоянии все понять, используя свою пассивную лексику, которая относится к пониманию терминов, даже если они не являются частью повседневного лексикона пациента).
- Если была создана универсальная версия языка, перевод становится менее очевидным, чем в любой индивидуальный подход, особенно если позднее потребуются перевод в еще одной стране, где говорят на том же языке. См. сценарий 2 для поиска вариантов, как справиться с этой ситуацией.

Факторы, которые необходимо учитывать, выбирая подход

Существует мнение, что ни один из подходов не лучше другого, так как каждый имеет свои преимущества и недостатки. Наша рекомендация — каждая ситуация и исследование должны оцениваться на основе рассмотрения каждого случая в отдельности. Мы разработали схему принятия решений для оказания помощи в этом процессе и обеспечения руководства в отношении случаев, в которых может быть оказано явное предпочтение одному подходу в сравнении с другими. Факторами, которые следует учитывать, являются участвующие языки и страны, а также содержание опросника и рекомендации разработчика.

Общие рекомендации

1. Решение о выборе наиболее подходящего подхода частично относится к культурным сходствам и различиям между странами, в которых говорят на целевых языках. Например, сходство часто будет больше между двумя южноамериканскими испаноговорящими странами, чем между Испанией и любой из стран Южной Америки. Лингвистическое сходство между французским во Франции и французским в Бельгии больше, чем между французским во Франции и французским в Канаде.
2. Некоторые вопросы нельзя перевести универсальным образом:
 - образовательные системы;
 - некоторая демография (например, уровень доходов и этническая принадлежность);
 - некоторое содержание деликатного или ограничительного характера может быть связано с культурой (например, сексуальное поведение или вопросы, относящиеся к алкоголю);

- использование ресурсов здравоохранения со ссылкой на конкретные учреждения (например, «Meals on Wheels» — «Еда на колесах»);
- государственные учреждения и программы, такие как медицинское страхование, инвалидность; и
- компьютерная терминология, в настоящее время требуемая для некоторых переводов ePRO. Соответствием этой меры методу является то, что всегда должно быть рассмотрено и никогда не предполагается.

Способы осуществления этих подходов к переводу

После принятия решения о подходе или комбинации подходов рекомендуются следующие общие процессы:

Метод адаптации одного языка

- Начинать стоит с существующей языковой версии.
- Два носителя языка (профессиональные переводчики или исследователи в ключевой стране) от целевой группы населения рассматривают существующий перевод и определяют условия использования, которые неприемлемы в целевой стране или будут там неправильно поняты.
- После сравнения комментариев рецензентов, рецензент-носитель языка (новый человек или один из упомянутых выше) подтверждает приемлемость пересмотренного документа в целевой стране или идентифицирует термины, которые могут там быть неправильно поняты.
- Выполнение резервного перевода пересмотренных элементов с целевого языка на английский для обеспечения точности.
- Внедрение изменений.
- Корректур пересмотренного инструмента.
- Когнитивное разбирательство с пациентами в каждой целевой стране для подтверждения, что новую версию хорошо понимают.

Универсальные рекомендации по переводу, применяемые к стандартной методологии перевода

- Основные переводчики должны быть происхождения из нескольких стран, особенно важно — из целевых для исследования. Например, для немецкого языка, один основной переводчик может быть из Германии, а другой — из Австрии. Если нужно включить много других стран, предполагается наличие более чем двух основных переводчиков.
- Синхронизатор должен быть знаком с использованием инструмента в разных странах, чтобы иметь возможность согласовать изменения таким образом, чтобы перевод не склонялся к той или иной стране.

- После обратного перевода несколько рецензентов или врачей вносят свой вклад в синхронизацию перевода и определяют проблемы для своей страны или региона.
- Переводчики работают вместе, чтобы найти решения, приемлемые для всех регионов. В случаях, когда такое решение невозможно, могут быть получены различные заключительные переводы, которые поддерживают большинство тех же формулировок, но требуется включение вариаций для конкретных стран.
- Корректировка (представителями разных стран, если поддерживается универсальная версия).
- Рекомендуется когнитивное разбирательство с пациентами в разных странах, чтобы подтвердить, что универсальный перевод понятен и приемлем. Для опроса пациентов необходимо достаточное количество людей из каждой целевой страны для получения обратной связи, в большинстве случаев, как минимум, пять человек от страны или региона. Если пациенты находят проблемы в универсальном переводе и вносят предложения об изменениях, может быть проведена та же адаптация универсальной версии языка на основе этой обратной связи с изменением только проблемных условий или предметов.

Метод перевода, специфический для каждой страны

Перевод нового инструмента (опросника), специфического для каждой страны, будет следовать руководству ISPOR по надлежащей практике [1], с переводчиками, синхронизаторами и представителями внутри страны от каждой страны, вовлеченной в процесс.

Заключение

При рассмотрении вопроса переводов для одного и того же языка, но в разных странах, ситуация довольно сложная. При недостатке литературы по этому вопросу, мы опирались на практический опыт. Мы исследовали несколько хорошо известных подходов: специфический подход для конкретной страны, подход языковой адаптации и универсальный подход. Разработка сценариев и схемы принятия решений подтвердила, что ни один из подходов не является лучшим или уместным в любой ситуации. Наша рекомендация — оценивать каждое исследование на индивидуальной основе. Кроме того, участвующие язык и страны и содержание инструментов должны быть оценены, чтобы определить наиболее уместный подход. Необходимы дополнительные исследования, чтобы сравнить результаты этих подходов. Например, одно эмпирическое исследование могло бы сравнить данные, собранные при универ-

сальном переводе в двух или более странах, используя различные трактовки для оценки, понимается ли один и тот же перевод по-разному в сравниваемых странах. Другим возможным исследованием будет сравнение данных, собранных в разных странах, использующих специфические версии для страны или адаптацию того же язык для оценки, показывают ли разные переводы существенные различия. Исследования, подобные этим, будут способствовать сбору эмпирических данных о различных подходах к переводу и обеспечат понимание методов оценки объединения данных из разных переводов.

РАЗДЕЛ 3 — Вопросы по сбору различных языковых и культурных данных

Введение

Растущее включение оценок исходов пациентами (PRO) в крупных международных исследованиях приводит к новым методологическим проблемам в анализе и интерпретации данных, полученных в ходе исследований. PRO часто разрабатываются на английском языке и переводятся на различные языки, необходимые для поддержки этих глобальных испытаний.

Существующая литература описывает стандартные методы лингвистической валидации для разработки высококачественных новых языковых версий PRO [1, 2]. Большинство этих методов, однако, сосредотачивается на качестве отдельных языковых версий для обеспечения семантической эквивалентности и не распространяется на решения о целесообразности объединения и анализа данных, полученных при использовании нескольких языковых версий.

Клинические исследования проводятся во многих странах, чтобы определить вариабильность данных и количество участников, включенных для оценки. Увеличение размера выборки имеет положительное влияние на силу и репрезентативность статистического анализа [8]. Однако, если данные неуместно объединяли, изначальные различия между разными источниками данных могут быть скрыты, уступая ошибочным результатам и выводам и, в конечном счете, поставляя недействительные данные [9].

Успешное объединение многонациональных наборов данных имеет решающее значение для достижения преимуществ, связанных с увеличенным размером выборки. Вопросы целесообразности объединения данных очень важны. Различия клинических данных между странами могут быть результатом различий в клинических стандартах оказания медицинской помощи, различия персонала сайтов, процедур и оценке смешанных случаев. Отличия культурных и языковых предпочтений в субъективном выражении результата представляют собой дополнительные проблемы для PRO.

В настоящее время не существует установленных критериев, касающихся дизайна исследования или аналитических требований для обеспечения уместности объединения данных, полученных из нескольких языков и культур. Предположение, что разные языковые версии одного и того же инструмента (если они были адаптированы с использованием соответствующей методологии) имеют эквивалентные психометрические свойства, а также выполняются одинаково в различных языковых группах, было использовано для поддержки объединения многонациональных предварительных данных [10]. Однако же, это непроверенные предположения. *Ло и др.* [11] предположили, что шаг по объединению данных может осуществляться только при уверенности, что различные языковые версии измерительных инструментов можно построить с той же метрикой.

Для объединения данных PRO из нескольких источников существуют разные виды эквивалентности. К ним относятся основные вопросы описательных характеристик населения, включая уровень грамотности, когнитивную эквивалентность понятий, используемых в документах сбора данных, эквивалентность в методах, используемых для вывода нескольких языковых версий, эквивалентность в способности инструментов продемонстрировать приемлемые психометрические свойства (психометрическую эквивалентность) и метрическую (измерение) эквивалентность в масштабах реагирования и озвучивания собранных данных.

Для объединения данных PRO полученных за время исследований в разных странах, первые шаги включают в себя обеспечение эквивалентности в области дизайна и методов исследования. Следующим шагом является обеспечение языковой и когнитивной эквивалентности концепций инструментов PRO (используя соответствующие методы языковой и культурной адаптации, см. раздел 2). Эти методы обсуждались в литературе [1, 2, 12]. Кроме того, эквивалентность измеряемых понятий может быть протестирована на фокус-группах или при индивидуальных интервью в новых странах/культурах, чтобы проверить действенность концепции.

Этот раздел содержит обсуждение методов, используемых для оценки измерения эквивалентности, которое должно быть рассмотрено в процессе объединения данных, полученных из различных языков и культур.

Эквивалентность измерений

Эквивалентность измерений представляет собой сложную концепцию с различными существующими определениями.

Хорн и Макардл [13] и *Маллен* [14] определяют измерение эквивалентности как отражение степени, в которой измерительный прибор и соответствующий

протокол сбора данных может принести надежные и достоверные данные о некоторых явлениях, представляющих интерес у различных групп населения. Эквивалентность измерений считается достигнутой, если в различных условиях проведения наблюдательных исследований процессы измерения дают одинаковые результаты [9].

Ло и др. [11] определяют измерение эквивалентности как различные языковые версии одного и того же документа, требующие схожих показателей элементов и уровня шкал с одинаковыми уровнями качества жизни, связанного со здоровьем (HRQoL), выраженных респондентами. *Дрэгзоу* (1984) также признает наличие измерительной эквивалентности, когда отношения между наблюдаемыми результатами тестов и измеренными показателями идентичны внутри по субпопуляций в базе данных [15]. Таким образом, основной целью измерений эквивалентности является ответ на вопрос, что является причиной различий в показателях между участвующими группами населения: действительная разница между этими группами в измеряемых показателях или другие причины, которые могут повлиять на объединенные данные.

Проще говоря, измерение эквивалентности — это аналитический процесс для оценки полученных свойств измерительных приборов, используемых в исследовании участвующих групп населения.

Методы демонстрации измерения эквивалентности

Для оценки измерения эквивалентности используются разнообразные методы, в том числе классическая теория тестирования, факторный анализ, моделирование структурных уравнений (SEM) и оценка различного функционирования заданий по отношению к различным группам испытуемых (DIF).

Классическая теория тестирования (СТТ)

Классическая теория тестирования основана на идее, что тест состоит из ряда элементов, каждый из которых — попытка измерить психологическую характеристику оценки. Каждая попытка измерить черту (пункт) производит отличный от предыдущей результат, потому что каждый тестируемый продукт не является идеально надежным. Классическая теория тестирования, при использовании в измерении эквивалентности, гарантирует, что каждая языковая версия имеет схожие теоретические значения (например, среднее значение, дисперсия и т.д.) в пределах подобной выборки. *Скотт-Леннокс и др.* [16] использовали классические элементы теории испытания надежности и валидности, чтобы проверить эквивалентность различных языковых версий MOS в популяции ВИЧ. Данные, относящиеся к MOS и SCL-57, были собраны с использованием пяти различных

языковых версий и дополнили данные, сообщенные самими пациентами, собранные у 363 ВИЧ-позитивных амбулаторных больных. Оценивалась внутри культурная и межкультурная психометрическая адекватность переводов, чтобы определить целесообразность перевода исследования медицинских результатов (MOS) ВИЧ. Внутренняя согласованность, дискриминация элементов и конвергенция элементов, а также эффекты крайних точек были оценены на разных языках. Авторы обнаружили, что, в общем, пять переводов имели подобные психометрические свойства в США на английском языке. *Шталь и др.* [17] исследовали поперечные и продольные корреляции между переводами опросника качества жизни больных астмой (AQLQ) и клинических оценок. Они пришли к выводу, что непротиворечивость поперечной корреляции между AQLQ и клинической шкалой по странам подтверждает достоверность переводов.

Факторный анализ

Факторный анализ опирается на понятие, что эквивалентность достигается, когда отношения между наблюдаемыми оценками и латентными переменными равны для обеих групп сравнения. Отсутствие эквивалентности может возникать по двум причинам: 1) значимые групповые различия для скрытой переменной, или 2) отсутствие эквивалентного измерения для всех групп. Первую причину легко проверить; вторая является более серьезной, поскольку это означает, что изменение функции между группами различно [16, 18—20]. *Штраус и Карпентер* [21] используют факторный анализ для оценки структуры факторов французского перевода пересмотренных *Строссом и Карпентером* критериев оценки шкалы результатов (SCOCSR). Они обнаружили, что структура фактора, так же, как и достоверность внутренней оценки и конвергентная валидность французского перевода, являются зеркальным отражением оригинальной английской версии шкалы.

Моделирование структурных уравнений (SEM). Это статистический метод для тестирования и оценки причинно-следственных связей за счет использования комбинации статистических данных и причинно-следственных предположений. SEM может быть использовано для оценки серии вложенных измерительных моделей [22], чтобы определить, присутствует ли измерение эквивалентности между языковыми группами.

Различное функционирование заданий по отношению к различным группам испытуемых (DIF). DIF происходит, когда люди из разных групп (например, языков/культур) с одним и тем же скрытым признаком имеют разную вероятность дать определенный ответ. Это особенность подхода теории отклика элемента (IRT), она была использована для оценки измерения эквивалентности между переводами. *Скотт и др.* [23]

сравнили 13 переводов европейской организации по исследованию и лечению рака (EORTC) опросника качества жизни -C30 (QLQ -C30) в 22 странах. Большинство языков показали аналогичные результаты с английским. Однако в каждом переводе был найден как минимум один случай значительной DIF. В другом исследовании, проведенном *Пагано и Gotay* [24], DIF был найден по нескольким пунктам EORTC QLQ-C30 в кавказской, филиппинской, гавайской и японской группах, хотя выбор метода использовался для того, чтобы оценить, от какого ряда факторов зависит измерение эквивалентности: возможно, самое главное — размер выборки из стран, представляющих интерес. Требования в отношении размеров выборки для анализа СТТ намного меньше, чем те, которые требуются для факторного анализа, SEM и DIF. Когда *Смит и др.* [18] использовали методы SEM, они обнаружили, что японский перевод шкалы суточного ритма не измеряет ту же смысловую величину, что и оригинальная английская версии шкалы. Однако, когда авторы использовали подход СТТ, они обнаружили эквивалентность между версиями.

Проблемы интерпретации

Независимо от того, какой метод используется для поддержки объединения многокультурных предварительных данных, нет существующих золотых стандартов для определения уровня принимаемого сходства или дисперсии. В то время как измерение эквивалентности может быть желаемой целью, должны быть разработаны практические решения и параметры для определения степени измерения эквивалентности, необходимой для поддержки объединения данных глобальных исследований. Необходимо дальнейшее изучение альтернативных решений этих вопросов.

Было бы полезным определить, является ли достижимым на практике установление уровня эквивалентности или нет, а также некоторые формы руководства по предложенным подходам и степень необходимого сходства. В идеале, это будет включать в себя тесты для изучения, могут ли быть объединены данные, и правила для принятия решения, когда данные PRO по стране должны быть исключены. *Смит и др.* [18] и *Харви и др.* [25] предполагают, что выборки должны быть как можно более близкими в начале исследования, чтобы факторы (конфаундинг) вмешивающиеся.

Тем не менее, априори можно относительно установить только то, что кросс-национальные различия переменных существуют. Кроме того, существует необходимость учитывать репрезентативность различных выборок в рамках многокультурных групп (*Харви и др.*, [25]). Как предполагалось ранее, кросс-культурная действительность понятий могла бы быть изучена при качественных исследованиях. Кроме того, это может быть изучено с помощью ана-

лиза существующих пробных выборок и обсуждено с местными клиническими экспертами.

Когда различия найдены, возможно, целесообразно будет провести качественное исследование в странах/культурах, представляющих интерес с группами пациентов и/или медицинских работников, чтобы определить, существуют ли какие-либо потенциальные объяснения различий. Если есть возможность, те же PRO могут быть использованы на других наборах данных.

Заключение

При оценке приемлемости объединения данных транснациональных исследований, свидетельства эквивалентности могут быть рассмотрены на различных уровнях, включая сходство дизайна (проекта) исследования и лингвистической/культурной эквивалентности PRO. Проводится дополнительное рассмотрение измерения эквивалентности. Есть несколько аналитических подходов, которые исследуют сходства и различия измерений эквивалентности внутри субпопуляций. Однако в настоящее время не существует конечных точек сходства или приемлемого количества различий, которые можно принять. Усилия по укреплению качества отдельных языковых версий по мере их разработки не могут обеспечить гарантию того, что сами меры будут обладать приемлемым уровнем сходства внутри популяций.

Исследователям рекомендуется применять широкий спектр количественных и качественных методов и анализ результатов этих исследований, чтобы определить, является ли тест адекватным для конкретной популяции [15].

Патрик и др. [26] предположили, что с целью проверки основных свойств измерений, таких как распределение ответов, внутренняя согласованность, надежность результатов повторного тестирования и действия на нескольких языках, необходимо рассмотреть характеристики населения и изменчивость данных, таких как распределение ответов, значимых стандартных отклонений и т.д., для различных языковых групп.

Исследователи призывают к усиленному сбору научных доказательств целесообразности многих принятых, но непроверенных способов лечения на многонациональных пробных данных [27]. Важно, что любые рекомендации, полученные из таких доказательств, должны быть как строго соответствующе для защиты качества будущих испытаний, так и достаточно гибкими, чтобы учесть быстро меняющийся характер и условия для клинических исследований, в которых все чаще используют PRO для оценки значимости фармакологических соединений.

Источник финансовой поддержки: никакой финансовой помощи для финансирования данной рукописи не предоставлялось.

Литература

1. Wild D., Grove A., Martin M.L., et al. ISPOR principles of good practice: the cross-cultural adaptation process for patient reported. *Value Health* 2005;8:94—104.
2. Acquadro C., Conway K., Hareendran A., Aaronson, N., for the ERIQA Group. Literature review of methods to translate health related quality of life questionnaires for use in multinational clinical trials. *Value Health* 2008;11:509—21.
3. Eremenco S., Cella D., Arnold B.J. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof* 2005;28:212—32.
4. Bonomi A.E., Cella D.F., Hahn E.A., et al. Multilingual translation of the functional assessment of cancer therapy (FACT) quality of life measurement system. *Qual Life Res* 1996;5:309—20.
5. Cella D., Hernandez L., Bonomi A.E., et al. Spanish language translation and initial validation of the functional assessment of cancer therapy quality-of-life instrument. *Med Care* 1998;36:1407—18.
6. Koller M., Aaronson N.K., Blazeby J., et al. Translation procedures for standardised quality of life questionnaires: The European Organisation for Research and Treatment of Cancer (EORTC) approach. *Eur J Cancer* 2007;43:1810—20.
7. Hunt S.M., Alonso J., Bucquet D., et al. Cross-cultural adaptation of health measures. *European Group for Health Management and Quality of Life Assessment. Health Policy* 1991;19:33—44.
8. Thumboo J., Fong K.Y., Chan S.P., et al. The equivalence of English and Chinese SF-36 versions in bilingual Singapore Chinese. *Qual Life Res* 2002;11:495—503.
9. Rungtusanatham M.J., Ng C.H., Zhao X., Lee T.S. Pooling data across transparently different groups of key informants: measurement equivalence and survey research. *Decis Sci* 2008;39:115—45.
10. Mark B.A., Wan T.T.H. Testing measurement equivalence in a patient satisfaction instrument. *West J Nurs Res* 2005;27:772—87.
11. Luo N., Chew L.H., Fong K.Y., et al. Do English and Chinese EQ-5D versions demonstrate measurement equivalence? An exploratory study. *Health Qual Life Outcomes* 2003;1:7.
12. Maneeriwongul W., Dixon J.K. Instrument translation process: a methods review. *J Adv Nurs* 2004;48:175—86.
13. Horn J., McArdle J. A practical and theoretical guide to measurement invariance in aging research. *Exp Aging Res* 1992;1:117—44.
14. Mullen MR. Diagnosing measurement equivalence in cross national research. *J Int Bus Stud* 1995;26:573—96.
15. Davies S., Little I.S., Ross R. Ensuring the measurement equivalence and appropriate use of personality assessments across cultures. Paper presented at the Annual Meeting of the Society for Industrial-Organizational Psychology, May 2006, Dallas.
16. Scott Lennox J.A., Wu A., Boyer G., Ware J.E. Reliability and validity of French, German, Italian, Dutch, and UK English translations of the Medical Outcomes Study HIV Health Survey. *Med Care* 1995.
17. Stahl E., Postma D.S., Juniper E.F., et al. Health-related quality of life in asthma studies. Can we combine data from different countries? *Pulm Pharmacol Ther* 2003;16:53—9.
18. Smith C.S., Tisak J., Bauman T., Green E. Psychometric equivalence of a translated circadian rhythm questionnaire: implications for ISPOR Task Force Translation Guidelines 439 between- and within- population assessments. *J Applied Psycho* 1991;76:628—36.
19. Wee H.L., Ravens-Sieberer U., Erhart M., Li S. Factor Structure of the Singapore English version of the KINDL Children Quality of Life Questionnaire. *Health Qual Life Outcomes* 2007.
20. Timmerman E.M., Hoogstraten J., Nauta M., Meijer K. Structural comparison of a translated dental attitude questionnaire: a factor analytic study. *Community Dent Oral Epidemiol* 1996;24:236—9.
21. Poirier S., Bureau V., Lehoux C., et al. A factor analysis of the Strauss and Carpenter revised outcome criteria scale: a validation of the French translation. *J Nerv Ment Dis* 2004;192:864—7.
22. Vandenberg R.J., Lance C.E. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods* 2000;4—69.3
23. Scott N.W., Fayers P.M., Bottomley A., et al. Comparing translations of the EORTC QLQ-C30 using differential item functioning. *Qual Life Res* 2006;1103—15.
24. Pagano I.S., Gotay C.C. Ethnic differential item functioning in the assessment of quality of life in cancer patients. *Health Qual Life Outcomes* 2005;3:60.
25. Harvey P.D., Fortuny L.A., Vester-Blackland E., De Smedt G. Cross national cognitive assessment in schizophrenia clinical trials: a feasibility study. *Schizophr Res* 2002;59:243—51.
26. Patrick D.L., Burke L.B., Powers J.H., et al. Patient-reported outcomes to support medical product labelling claims: FDA perspective. *Value Health* 2007;10(Suppl.):S125—37.
27. Lenderking W.R. Comments on the ISPOR Task Force Report on Translation and Adaptation of Outcomes Measures: guidelines and the need for more research. *Value Health* 2005;8:92—3.

Инструменты для сбора данных сообщений об исходах от медицинского вмешательства от пациентов (PRO) применяемые в педиатрии и предназначенные для инструкции для медицинского продукта: Отчёт Исследовательской группы ISPOR по разработке Надлежащих Исследовательских Практик о сообщаемых пациентами исходах

Louis S. Matza, PhD¹, Donald L. Patrick, PhD, MSPH², Anne W. Riley, PhD, MS³,
John J. Alexander, MD, MPH⁴, Luis Rajmil, MD, PhD, MPH⁵, Andreas M. Pleil, PhD⁶,
Monika Bullinger, PhD⁷

¹ — Outcomes Research, United BioSource Corporation, Bethesda, MD, USA

² — Department of Health Services, University of Washington, Seattle, WA, USA

³ — Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

⁴ — Division of Anti-Infective Products, FDA, Silver Spring, MD, USA;

⁵ — IMIM-Hospital del Mar Medical Research Institute, and Catalan Agency for Health Information, Assessment and Quality, Barcelona, Spain

⁶ — Pfizer Global Pharmaceuticals, Pfizer, Inc., San Diego, CA, USA

⁷ — Department of Medical Psychology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany

Переводчики:

Гапешин Роман Андреевич — Медицинский факультет, Санкт-Петербургский государственный университет, Россия
Мальшев Станислав Михайлович — Медицинский факультет, Санкт-Петербургский государственный университет,

Россия

Рецензент:

Колбин Алексей Сергеевич — д.м.н., профессор, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Резюме. *Общие сведения:* Инструменты для сбора данных сообщений об исходах от медицинского вмешательства от пациентов (PRO) у детей и подростков часто включают в клинические исследования с целью сбора данных, подтверждающие сведения, изложенные в инструкции по применению медицинского продукта. *Цель:* Целью настоящего отчёта исследовательской группы было создание рекомендаций по надлежащей практике для исследований PRO в педиатрии,

организованных с целью предоставления данных для принятия управленческих решений и подтверждающих сведений, изложенных в инструкции по применению медицинского продукта. Рекомендации основываются на консенсусе междисциплинарной группы исследователей, объединённых в исследовательскую группу, ассоциированную с Международным Обществом Фармакоэкономики и Исследования Исходов (International Society for Pharmacoeconomics and Outcomes Research (ISPOR)). В тех случаях, когда был недостаток доступных сведений или, когда общие принципы неприменимы ко всем ситуациям, настоящая исследовательская группа указывает факторы, которые следует рассмотреть при принятии решений касательно дизайна и использования педиатрических инструментов PRO, при этом обозначая проблемы, требующие дальнейших исследований. *Надлежащие Исследовательские Практики:* Обсуждены пять надлежащих исследовательских практик: 1). Рассмотреть особенности развития и определить критерии для введения PRO, основанные на возрасте пациента: опираясь на предыдущие исследования, проанализированы четыре возрастные группы (<5 лет, 5-7 лет, 8-11 лет, 12-18 лет). Эти возрастные группы рекомендуют в качестве отправной точки при принятии решений. Конкретные возрастные рамки следует определять отдельно для каждой популяции и инструмента PRO; 2). Установить содержательную валидность педиатрических инструментов PRO: в данном разделе обсуждают преимущества привлечения детей в качестве экспертов, а также стратегии выявления концепций и когнитивных интервью с детьми; 3). Определить, необходим ли инструмент PRO: обсуждают отличия между двумя типами (доверительный и наблюдательный) опросников, заполняемых опрашиваемыми, и даны рекомендации; 4). Убедиться, что инструмент создан и оформлен соответственно целевой возрастной группе. К рассмотренным факторам относят лексику, связанный со здоровьем, уровень чтения, шкалы ответов, память, длительность, наглядность, оформление, организационные подходы и электронный сбор данных; 5). Рассмотреть кросс культурные аспекты. Выводы: Для формирования методического руководства для будущих исследований, особенно включающих детей младшего возраста и сообщения от родителей, требуются дополнительные данные. Поскольку, данные PRO всё чаще используются для подтверждения и улучшения информации в инструкции к медицинским продуктам в педиатрии, стоит ожидать новую информацию, касающуюся стандартов, этих инструментов. Использование инструментов PRO в клинических исследованиях и при оформлении заявки на регистрацию медицинского продукта помогут убедиться в том, что переживание ребёнком болезни и процесса лечения адекватным образом отображены и проанализированы при принятии административных решений.

Ключевые слова: подростки, дети, ISPOR, маркировка медицинского продукта, сообщаемые пациентами исходы, педиатрия, PRO, исследовательская группа

Pediatric Patient-Reported Outcome Instruments for Research to Support Medical Product Labeling: Report of the ISPOR Good Research Practices for the Assessment of Patient-Reported Outcomes in Children and Adolescents Task Force

Abstract. Patient-reported outcome (PRO) instruments for children and adolescents are often included in clinical trials with the intention of collecting data to support claims in a medical product label. The purpose of the current task force report is to recommend good practices for pediatric PRO research that is conducted to inform regulatory decision making and support claims made in medical product labeling. The recommendations are based on the consensus of an interdisciplinary group of researchers who were assembled for a task force associated with the International Society for Pharmacoeconomics and Outcomes Research (ISPOR). In those areas in which supporting evidence is limited or in which general principles may not apply to every situation, this task force report identifies factors to consider when making decisions about the design and use of pediatric PRO instruments, while highlighting issues that require further research. Five good research practices are discussed: 1) Consider developmental differences and determine age-based criteria for PRO administration: Four age groups are discussed on the basis of previous research (<5 years old, 5-7 years, 8-11 years, and 12-18 years). These age groups are recommended as a starting point when making decisions, but they will not fit all PRO instruments or the developmental stage of every child. Specific age ranges should be determined individually for each population and PRO instrument. 2) Establish content validity of pediatric PRO instruments: This section discusses the advantages of using children as content experts, as well as strategies for concept elicitation and cognitive interviews with children. 3) Determine whether an informant-reported outcome instrument is necessary: The distinction between two types of informant-reported measures (proxy vs. observational) is discussed, and recommendations are provided. 4) Ensure that the instrument is designed and formatted appropriately for the target age group. Factors to consider include health-related vocabulary, reading level, response scales, recall period, length of instrument, pictorial representations, formatting details, administration approaches, and electronic data collection (ePRO). 5) Consider cross-cultural issues. Additional research is needed to provide methodological guidance for future studies, especially for studies involving young children and parents' observational reports. As PRO data are increasingly used to support pediatric labeling claims, there will be more information regarding the standards by which these instruments will be judged. The use of PRO instruments in clinical trials and regulatory submissions will help ensure that children's experience of disease and treatment are accurately represented and considered in regulatory decisions.

Keywords: adolescents, children, ISPOR, medical product labeling, patient-reported outcomes, pediatrics, PRO, task force

Об Исследовательской Группе

В марте 2009 года Совет по Стратегии в области Науки о Здоровье ISPOR (ISPOR Health Science Policy Council) реко-

мендовал Совету Директоров ISPOR собрать Исследовательскую группу ISPOR по разработке Надлежащих Исследовательских практик о сообщаемых пациентами исходах (PRO) (ISPOR Good Research Practices Patient-Reported Outcomes Task Force), чтобы целенаправленно изучить оценку сообщаемых пациентами исходов у детей и подростков. Совет Директоров одобрил формирование Исследовательской группы по PRO в марте 2009 года. Председатель (доктор Матза) и сопредседатель (доктор Патрик) Педиатрической Исследовательской Группы по PRO выбрали членов исследовательской группы, исходя из их опыта в оценке PRO и исследованиях, касающихся детей и подростков. Участники были отобраны таким образом, чтобы был представлен широкий спектр мнений, в том числе правительства (Управление по санитарному надзору за качеством пищевых продуктов и медикаментов Соединённых Штатов Америки), научного сообщества, исследовательских организаций и фармацевтической промышленности. Кроме того, исследовательская группа характеризовалась международным участием, в её состав входили представители Германии, Испании и Соединённых Штатов Америки.

Первоначально Исследовательская Группа собиралась примерно каждые два месяца с помощью телеконференций, чтобы разработать основные положения и обсудить темы, которые должны войти в отчёт. Личные встречи прошли в Новом Орлеане в Октябре 2009 года и в Атланте в мае 2010 года, где эти темы обсуждались далее и был достигнут консенсус по части рекомендаций. Кроме того, председатель исследовательской группы неоднократно проводил индивидуальные телеконференции с членами группы, участвовавшими в формировании черновой версии отчёта. Всеми членами группы было просмотрено множество вариантов отчёта и регулярно предоставлялись отзывы на них в устной и письменной форме.

Предварительные выводы и рекомендации были представлены на форуме в рамках 15-й Ежегодной Международной Конференции ISPOR в мае 2010 года. Дополненные выводы и рекомендации были представлены на форуме в рамках 17-й Ежегодной Международной Конференции ISPOR в июне 2012 года. Комментарии, полученные в ходе этих двух форумов, были учтены в последующих версиях отчёта.

Черновая версия данного отчёта была предоставлена Обзорной Группе ISPOR (включавшей более 400 участников) в марте 2012 года. Проверенный черновик был предоставлен всему сообществу ISPOR в январе 2013 года. В ходе этих двух этапов обзора было получено свыше 250 письменных комментариев от 40 членов ISPOR. Письменные комментарии также пересылались участниками различных регуляторных и финансовых органов, в том числе тремя рецензентами из Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA), одному из Института качества и эффективности в здравоохранении Германии (Institute for Quality and Efficiency in Health Care (IQWiG)) и представителем французского Национального бюро здоровья (Haute Autorité de Santé [HAS]) и Европейской сети по оценке технологий здравоохранения (EUnetHTA).

Все комментарии были рассмотрены, большинство из них были существенные и конструктивные. В дальнейшем, комментарии обсуждали исследовательской группой в ходе нескольких телеконференций и соответствующим образом учитывали в рецензированных версиях отчёта. Когда все участники исследовательской группы достигли согласия, окончательная версия отчёта была опубликована в «Value in Health» в апреле 2013 года.

Все письменные комментарии опубликованы на веб-сайте ISPOR на веб-странице исследовательской группы: <http://www.ispor.org/TaskForces/PROChildrenAdolescents.asp>. Отчёт и веб-страница исследовательской группы также доступны на официальном сайте ISPOR (www.ispor.org) через фиолетовое меню Инструментов для Исследования, Надлежащие Практики для Исследования Исходов, рубрика: Методы изучения исходов, ориентированных на пациента и описанные клиницистом, ссылка: Оценка PRO у детей и подростков. Также на веб-странице исследовательской группы доступен список рецензентов.

Введение

Инструменты, с помощью которых получают информацию, сообщаемых пациентами исходов (PRO) подразумевает сбор сведений о состоянии здоровья напрямую от пациента без интерпретации его мнения клиницистом, исследователем или кем-либо ещё [1, 2]. Многие аспекты заболеваний известны только самим пациентам, и прямая оценка точки зрения необходима для чёткого понимания переживания больным заболевания и процесса лечения. В последние годы стали подчёркивать важность систематической разработки и валидации инструментов PRO для использования в клинических исследованиях, оценивающих эффективность медицинского продукта. Инструменты PRO часто включают в клинические

исследования с целью сбора данных, подкрепляющих сведения о медицинском продукте в инструкции по его применению [3, 4].

Управление по санитарному надзору за качеством пищевых продуктов и медикаментов США (The US Food and Drug Administration (FDA)) и Европейское агентство лекарственных средств (European Medicines Agency (EMA)) опубликовали руководство по оценке PRO. Руководство FDA, опубликованное в предварительном варианте в 2006 году и завершённое в 2009 году, оказало существенно влияние на финансируемые фармацевтической индустрией исследования PRO [1]. Данное руководство содержит обзор использования PRO в контексте разработки медицинских продуктов, а также рекомендации по созданию и оценке таких методов. Неболь-

шой раздел руководства посвящен инструментам PRO, пригодным для использования в отношении детей и подростков (Раздел III.G.1). Он начинается с утверждения, что «проблемы, связанные с процессом разработки педиатрических инструментов PRO те же, что и обозначенные для взрослых». Далее, в разделе отмечается, что использование инструментов PRO в педиатрических популяциях ставит особые задачи, не возникающие при исследовании PRO у взрослых. Упомянуто несколько задач, в том числе соответствующий возрасту лексикон, понимание концепции здоровья, необходимость определения нижнего возрастного порога, начиная с которого дети могут давать надёжные и ценные описания, и соответствующее использование ответов, данных не самими пациентами, а другими информантами.

Как и FDA, EMA дала рекомендации для изучения PRO, особенно в отношении оценки качества жизни, связанного с состоянием здоровья (health-related quality of life (HRQOL)) [5]. Руководство EMA, тем не менее, не рассматривает применение инструментов PRO для детей и подростков. В целом, существует лишь ограниченное количество доступных руководств по исследованиям, включающим оценку PRO в педиатрии при разработке медицинских продуктов.

Таким образом, целью настоящего отчёта исследовательской группы было рекомендовать надлежащие практики для исследований PRO в педиатрии, направленных на предоставление информации для принятия управленческих движений и поддержку сведений в инструкции к медицинскому продукту. Рекомендации в данном отчёте основаны на консенсусе междисциплинарной исследовательской группы, ассоциированной с Международным Обществом по Фармакоэкономике и Исследованию Исходов (International Society for Pharmacoeconomics and Outcomes Research (ISPOR)). Надлежащие исследовательские практики суммированы в табл. 1.

Проблемы, связанные с выбором, разработкой и применением инструментов PRO для детей и подростков, были изложены и проанализированы во многих публикациях [6—22]. Несколько статей и глав книг также содержали списки и разборы универсальных и специфичных для того или иного заболевания инструментов PRO для детей и подростков [6, 9, 14, 23—28]. Настоящий отчёт исследовательской группы отличается от этих работ сосредоточенностью на инструментах PRO в педиатрии в контексте разработки и регистрации медицинских продуктов.

Рекомендации в этом отчёте основаны, насколько это было возможно, на опубликованных исследованиях. Оценка PRO в педиатрии, тем не менее, является развивающимся направлением исследований, и эмпирические данные ограничены для некоторых аспектов дизайна, разработки, валидации и применения инструментов. Таким образом, на сегодняшний день невозможно дать чёткие рекомендации по

некоторым проблемам, изложенным в настоящем отчёте. В таких ситуациях в отчёте исследовательской группы обсуждается дизайн и использование инструментов PRO для детей и подростков. Кроме того, данный отчёт подчёркивает те сферы, в которых для развития оценки PRO в педиатрии необходимы дальнейшие исследования.

Педиатрическая литература, посвящённая PRO, часто уделяет внимание вопросам особенностям развития и возрастным ограничениям [16, 21, 29, 30]. Например, в каком возрасте дети впервые могут описать состояние своего здоровья, и когда дети могут работать с более абстрактными понятиями? Наша исследовательская группа не считает, что возможно ввести возрастные рамки, которые можно будет использовать во всех ситуациях. Для каждого отдельного исследования или измерения PRO ответ на все эти вопросы о возрасте будут зависеть от множества факторов, в том числе от оцениваемой концепции, формата инструмента, степени независимости, требующейся от респондентов (например, анкеты для самостоятельного заполнения либо измерения, осуществляемые интервьюером) и характеристик интересующей популяции (например, медицинское или психиатрическое состояние). Определение возрастных ограничений также осложняется существенной вариабельностью когнитивного уровня в пределах любой отдельно взятой возрастной группы, а также возможными половыми различиями в восприятии и поведении в условиях заболевания [31].

Несмотря на перечисленные проблемы, представляется возможным предложить рекомендации, касающиеся способностей детей в различных возрастных группах на основании исследований в области психометрии и когнитивного развития. В то время как не существует возрастных ограничений, подходящих для всех инструментов PRO, необходимо обсудить общие возрастные рамки, чтобы суммировать данные научной литературы и выработать практические рекомендации. Исследовательская группа считает, что педиатрические выборки можно разделить на четыре общие возрастные группы с точки зрения оценки PRO для оценки медицинских продуктов, но эти возрастные группы следует использовать лишь в качестве отправной точки для принятия решений. Конкретные возрастные рамки должны быть определены индивидуально для каждого инструмента PRO, и возрастные группы для инструмента PRO следует испытывать при помощи когнитивных интервью всякий раз, когда инструмент начинает использоваться для новой целевой популяции.

Кроме того, как уже упоминалось, дети в пределах каждой возрастной группы, скорее всего, будут различаться по способности достоверно проходить опросники. Исследователям потребуется иметь представление об этих различиях на этапе сбора данных. В то время как некоторые исследования полагаются

Надлежащие исследовательские практики, рассмотренные в данном отчёте исследовательской группы PRO

Надлежащие Исследовательские Практики	Комментарии и Рекомендации
Рассмотреть особенности развития и определить критерии для введения PRO, основанные на возрасте пациента	Были проанализированы четыре возрастные группы. Их следует использовать в качестве отправной точки для принятия решений. Определить возрастные диапазоны для каждой ситуации невозможно. Конкретные возрастные диапазоны определяются индивидуально для каждого инструмента PRO и проверяются с помощью когнитивных интервью в каждой новой целевой популяции. Младше 5 лет: Отсутствие чёткого подтверждения надёжности или валидности показателей сообщений от детей. 5-7 лет: Сбор сообщений от детей возможен, но надёжность и валидность зачастую сомнительны. 8-11 лет: Повышение надёжности и валидности сообщений от детей 12-18 лет: Собственные сообщения оказываются предпочтительными.
Установить валидность содержания инструментов PRO в педиатрии	Дети и подростки могут быть эффективными экспертами по материалам. В большинстве случаев, детей следует включать в качественные исследования, направленные на установление содержательной валидности педиатрических PRO. С предполагаемым респондентом надлежит проводить когнитивные интервью. Детей следует интервьюировать по поводу сообщений от детей, а родителей — по поводу сообщений от родителей. Содержательную валидность следует демонстрировать для узких возрастных подгрупп.
Определить, необходим ли инструмент сообщённых информантом исходов	Описанные информантом исходы включают доверительные и наблюдаемые опросники. Когда дети в целевой возрастной группе способны самостоятельно воспользоваться инструментом PRO, следует использовать данные сообщений от детей. Во-вторых, когда дети в целевой возрастной группе не способны пройти опросник PRO, следует использовать данные сообщений от информантов. Данные сообщений от информантов, насколько это возможно, должны оценивать наблюдаемые материалы.
Убедиться, что инструмент создан и оформлен соответственно целевой возрастной группе	Лексикон и уровень чтения с учётом состояния здоровья Шкала ответа Период воспроизведения Длительность инструмента Графические изображения Оформление Административные аспекты Сбор электронных данных (эPRO)
Рассмотреть кросс-культурные проблемы	Содержательная валидность и подход к изучению педиатрического инструмента PRO нужно будет перепроверять в условиях каждой новой культуры.

Надлежащие исследовательские практики по разработке и применению педиатрических инструментов PRO

Надлежащая исследовательская практика 1: Рассмотреть особенности развития и определить критерии для использования PRO, основанные на возрасте пациента

на мнение интервьюеров или сборщиков данных, более стандартизированный подход подразумевает изучение способности детей участвовать в оценке PRO. Например, недавнее исследование оценивало уровень чтения у каждого ребёнка с помощью Индикатора Уровня Чтения от American Guidance Service (AGS) [32]. Детям, демонстрировавшим способность к чтению не ниже уровня четвёртого класса, предлагали заполнить анкету письменно. Дети, не достигшие этого уровня, проходили интервью. Этот скрининговый подход основывается на идее о том, что каждое умение ребёнка, а не собственно возраст, должен определять возможность и способ измерения PRO. Наша исследовательская группа считает крайне важным дальнейшее исследование методов скрининга способностей к прохождению инстру-

ментов PRO и путей решения проблемы их вариативности в возрастных группах.

Также следует отметить, что общепринятых определений детства и подросткового возраста не существует. Определения различаются в зависимости от культуры и законодательства. Определение, принятое Всемирной Организацией Здравоохранения [33], впервые появилось в Статье 1 Конвенции Организации Объединённых Наций о правах ребёнка: «Ребёнком является каждое человеческое существо до достижения 18-летнего возраста» [34]. Центры по контролю и профилактике заболеваний цитировали это определение по крайней мере один раз [35], но другой документ Центров по контролю и профилактике заболеваний касался более широкой группы лиц до 19 лет [36]. Старшую часть этого диапазона,

начиная приблизительно с 12-13 лет, обычно называют «подростковым возрастом» [37, 38]. Признавая нечёткость и изменчивость данных определений, наша исследовательская группа работает в первую очередь в возрастном диапазоне от 0 до 18 лет. В настоящем отчёте термин «подростковый возраст» относится к лицам строго от 12 до 18 лет, в то время как термин «младший возраст» в целом касается детей до 8 лет.

Возрастная группа 1 (младше 5 лет)

Первая возрастная категория начинается с момента рождения и включает новорождённых и детей дошкольного и дошкольного возраста. Хотя для детей младше 5 лет разработаны инструменты для самостоятельно сообщаемых исходов, чётких данных о надёжности или валидности изучения таких описаний получено не было [39]. Таким образом, оценка состояния здоровья в самой младшей возрастной группе должна основываться на клиническом обследовании и обсервационных сообщениях родителей или других взрослых.

Возрастная группа 2 (5-7 лет)

Вторая возрастная категория включает детей младшего возраста от 5 до 7 лет. Детям этого возрастного диапазона доступны общие и специфичные для отдельных заболеваний методы самостоятельных сообщений для оценки состояния здоровья, в том числе симптомов и HRQOL [9, 10, 14, 24, 27, 28, 40]. Некоторые исследователи высказываются за прямую оценку точки зрения ребёнка в этой возрастной группе [8, 30, 41], существует несколько статей с доступными психометрическими данными уже в пятилетнем возрасте. Например, дети пяти лет, прошедшие Педиатрический опросник качества жизни (Pediatric Quality of Life Questionnaire (PedsQL)), широко используемый для общего измерения HRQOL, дали минимальное количество нерелевантных результатов (2,8%) и надёжную внутреннюю непротиворечивость с суммарной альфой Кронбаха 0,86 для всех 23-х параметров [21]. Тем не менее, эти данные не подтверждаются другими исследованиями с куда менее обнадеживающими психометрическими результатами. При валидации версии с сообщениями детей Профиля детского здоровья и болезни (Child Health and Illness Profile (CHIP-SE)) надёжность внутренней непротиворечивости и воспроизводимости результатов для детей от 6 до 7 лет оказалась ниже общепринятых пределов (то есть <0,70) для трёх или двух подшкал (альфы Кронбаха для множественных подшкал варьировались от 0,64 до 0,83; коэффициенты внутригрупповой корреляции [КВК] = 0,35-0,69), однако статистика надёжности была значительно лучше в группе 8-11 лет (альфы Кронбаха =

0,72-0,85; КВК = 0,71-0,80) [42]. Аналогично, разработчики процедур самостоятельных сообщений для детей с эпилепсией указали на повышение надёжности воспроизводимости с 6-7 лет (КВК = 0,18-0,52) до 8-15 лет (КВК = 0,59-0,69) [43]. Более того, обзор 53-х опросников самостоятельных сообщений, связанных со здоровьем, у детей до 8 лет указывает на то, что только 51% опросников соответствовал минимальным стандартам надёжности внутренней непротиворечивости (альфа Кронбаха = 0,70) и лишь 23% опросников соответствовали стандартам надёжности воспроизводимости (КВК = 0,70) [24]. Основанная на доступных данных, полученных от детей младшего возраста, надёжность воспроизводимости представляется особенно проблематичным параметром, в то время как надёжность внутренней непротиворечивости обычно оказывается ближе к общепринятым стандартам.

В добавок к сомнениям в надёжности, некоторые данные указывают на то, что многие дети в возрасте от 5 до 6 лет, возможно, не пользуются все пять пунктов обычной шкалы Лайкерта. Напротив, они больше доверяют вариантам в середине или в конечных точках шкалы, как если бы она состояла из трёх пунктов [19, 44]. У детей этой возрастной группы также может быть ограничен лексикон, связанный со здоровьем. Исследование с использованием когнитивных интервью показало, что понимание детьми терминов, широко используемых в опросниках состояния здоровья (например, «нервный», «энергия» и «боль»), обычно ограничен в возрасте 5 лет, но затем постепенно расширяется с 6 до 8 лет (рис. 2) [19].

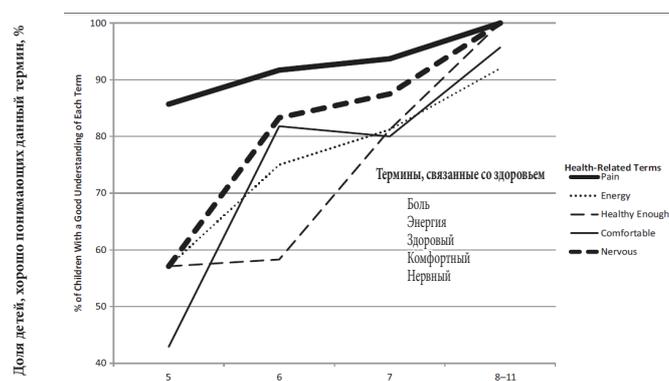


Рис. 2. Возрастные различия в понимании ребёнком терминов, связанных со здоровьем

Данные из Qual Life Res, 10, *Rebok G, Riley A, Forrest C, et al.*, Elementary school-aged children's reports of their health: a cognitive interviewing study, 59–70, 2001. С любезного разрешения Springer Science and Business Media.

Исследования возможностей когнитивной обработки указывает на то, что данные PRO, предоставленные детьми в этом возрастном диапазоне, могут быть полезны для оценки некоторых концепций, но должны интерпретироваться с осторожностью [29,

45]. Дети в возрасте от 5 до 7 лет, как правило, понимают, что значит чувствовать себя здоровым/нездоровым или хорошо/плохо, и обычно способны применить это знание к самим себе, часто с некоторыми особенностями. Также предполагают, однако, что дети в этом возрастном диапазоне могут иметь неравные когнитивные возможности (то есть выраженные способности в одном когнитивном домене, например, в распознавании слов, но слабые в другом, например, во внимании). Более того, некоторые дети могут быть медлительными, неумелыми или склонными к ошибкам при выборе варианта ответа, наилучшим образом выражающего их точку зрения [29, 45].

В целом, хотя анкеты самостоятельного сообщения могут быть полезны для сбора информации о здоровье детей в возрасте от 5 до 7 лет, накопленные данные позволяют усомниться в надёжности и точности их ответов. Для инструментов PRO, которые предполагают использовать в принятии решений, им нужно продемонстрировать валидность материалов и соответствующие психометрические свойства в целевом возрастном диапазоне, что может быть затруднительно в этой возрастной группе. Вероятность выявления приемлемых свойств измерения для детей в возрасте от 5 до 7 лет может быть повышена путём использования или создания анкет PRO, оценивающих простые концепции с понятным изложением, лексиконом, соответствующим возрасту, и меньшим числом вариантов ответа по сравнению с опросниками для старших детей. Если в данной возрастной группе продемонстрировать адекватные свойства анкеты невозможно, можно воспользоваться анкетами исходов, описанных информантом, заполняемыми родителями или другими взрослыми (см. Надлежащая Клиническая Практика 3 в этом отчёте исследовательской группы). Мы призываем к дальнейшему исследованию процедур PRO, выполняемых детьми, в надежде, что опыт и точка зрения детей может быть рассмотрена в ходе принятия решений.

Возрастная группа 3 (8-11 лет)

Третья возрастная группа состоит из старших детей, примерно от 8 до 11 лет. Способность детей независимо проходить анкетирование о состоянии здоровья, как считают, в этом возрастном диапазоне расширяется. Как было указано ранее в обсуждении возрастной группы 2, в ходе нескольких исследований среди детей в возрасте от 8 лет и старше была выявлена повышенная по сравнению с младшими группами надёжность, понимание слов, связанных со здоровьем, владение шкалами Лайкерта [19, 42, 43]. Самостоятельные сообщения детей в старшем возрастном диапазоне нередко соответствуют общепринятым стандартам надёжности универсальных и специфичных для заболевания методов измерения [21, 27, 46].

Результаты недавнего исследования, основанного на использовании когнитивных интервью и проведённого группой педиатров из Patient Reported Outcomes Measurement Information System (PROMIS), указали на наличие соответствующих умений у детей в этой возрастной группе [47]. Выборка в этом качественном исследовании, опиравшемся на педиатрический круг понятий, включала 24 ребёнка в возрасте 8-9 лет. Подобно подросткам, дети в этой возрастной группе были способны воспринять большинство понятий, вариантов ответа, инструкций и время воспроизведения при встрече с трудными для понимания словами.

Изучение умственного развития детей также предоставляет данные в поддержку методологии самостоятельных сообщений для детей, начиная с диапазона от 8 до 11 лет [29]. На этом этапе развития у детей формируются когнитивные способности, необходимые для прохождения опросников о здоровье. К ним относят: 1). понимание концепций здоровья; 2). способность к самоконтролю, в том числе к поддержанию внимания, минимизации недовольства и избеганию отвлекающих факторов; 3). способности к когнитивной обработке, в том числе к пониманию задания, удержанию вопроса в рабочей памяти в течение времени, достаточного для дачи ответа, оценке применимости вопроса к респонденту, оценке различий между вариантами ответа и выбору оптимального варианта [29].

В целом, проводят всё больше и больше исследований, подтверждающих осуществимость и надёжность самостоятельных сообщений от детей в возрасте 8-11 лет. Таким образом, оказывается возможным использование данных самостоятельных сообщений от этой возрастной группы в исследованиях, направленных на оценку медицинского продукта. Как и для всех остальных контрольных точек PRO, необходим тщательный качественный анализ, чтобы убедительно показать способность детей этой возрастной группы к пониманию инструментов PRO.

Возрастная группа 4 (12-18 лет)

Четвёртая возрастная группа включает в себя подростков примерно от 12 лет. Большинство подростковых PRO разработаны и валидированы для респондентов в возрасте 16, 17 или 18 лет [27]. Учитывая, что инструменты самостоятельных сообщений для подростков соответствуют общепринятым психометрическим стандартам, их использование не должно быть сопряжено с особыми сложностями [21, 28, 46]. Тем не менее, основная проблема в разработке инструментов PRO для подростков заключается в фиксации релевантного данной возрастной группе контента. Хотя элементы оценки симптомов могут быть теми же, что используются для взрослых, разработчикам инструментов для оценки влияния

симптома или функционального статуса следует принимать во внимание уникальные социальные и эмоциональные аспекты данного периода развития. Это было показано в качественном исследовании, проводившегося с использованием фокус-групп и интервьюирования подростков 12-18 лет, целью которого был сбор концепций инструмента для измерения качества жизни [48]. Полученная концептуальная модель выдвигает на первый план уникальность подросткового возраста, с широким спектром соответствующих концепций: поддержка взрослыми, свобода, взаимоотношения со сверстниками, самоидентификация, духовность, удовлетворённость жизнью, финансовые средства, видение будущего и образование. Многие из этих областей более сложны, нежели концепции, которые могут включаться в инструменты для детей до 12 лет. Кроме того, такие сферы, как взаимоотношения со сверстниками, видение будущего и образование могут включать контент, отличающийся от используемого в инструментах PRO у взрослых. Таким образом, многие материалы PRO вряд ли подойдут подросткам. Определение подходящего состава элементов для данной возрастной группы осложняет гетерогенность в этом возрастном диапазоне. Например, контент, релевантный для 18-летнего, не обязательно релевантен для 12-летнего. В целом, материалы самостоятельных сообщений у подростков зачастую указывают на психометрические качества, необходимые в качестве ключевой контрольной точки в клинических исследованиях. В то же время, качественным исследованиям также потребуется продемонстрировать, что конкретный инструмент обладает содержательной валидностью, специфичной для целевой возрастной группы.

Лица старше 18 лет в целом проходят PRO, разработанные для использования во всех взрослых возрастных группах. Однако может оказаться полезным анализ релевантности инструментов, ранее валидированных для молодых взрослых, обладающих собственными характеристиками. Исследователи предполагают, что возраст от 20 до 29 лет может рассматриваться как «ранняя взрослость» [49, 50]. Этот этап жизни, по-видимому, связан с задачами, особенными с точки зрения уровня независимости, личности, карьерного роста и взаимоотношений. Исследования, касающиеся развития и адаптации на данном этапе, будут публиковаться в новом научном журнале «Emerging Adulthood», первый выпуск которого ожидается в 2013 году. Некоторые процедуры PRO для взрослых могут потребовать качественного анализа с целью подтверждения содержательной валидности всех элементов для данной переходной возрастной группы.

В целом, при разработке новых инструментов PRO или использовании существующих инструментов для исследований, включающих оценку медицинского продукта, необходимо продемонстриро-

вать возрастную целесообразность инструмента для целевой популяции. Возрастная целесообразность PRO должна быть обоснована совокупностью качественных и количественных исследований. В первую очередь следует провести когнитивные интервью, чтобы проверить, все ли аспекты инструмента подходят для целевой возрастной группы. Эти интервью должны оценить мнение респондента об уровне словаря, содержанием элементов, периоде воспроизведения, вариантах ответов, инструкциях, понятности, релевантности и чёткости формулировок. Такое качественное исследование следует проводить на выборке адекватного объёма на верхней и нижней границе целевого возрастного диапазона, чтобы показать, что имеется достаточно данных для установления выбранных возрастных рамок. Далее может быть проведён качественный анализ для оценки психометрических характеристик инструмента для выборки, соответствующей заданному возрастному диапазону детей в запланированных клинических исследованиях.

Надлежащая исследовательская практика 2: установить содержательную валидность педиатрических инструментов PRO

Дети как эксперты содержания

В последние годы установление содержательной части подчёркивается в качестве критического этапа разработки и валидации инструментов PRO [1, 51, 52]. Существуют различные определения содержательной валидности [53], но определение, предложенное предыдущей исследовательской группой ISPOR, представляется соответствующим большинству источников, в том числе руководству FDA: «Степень, с которой инструмент содержит релевантные и значимые аспекты изучаемой с его помощью концепции» [54]. Также достигнуто согласие в том, что содержательная валидность устанавливается в первую очередь посредством качественного анализа, включающего прямые входные данные от целевой популяции.

Подобное качественное исследование, детально описанное ранее, принципиально проводить в два последовательных этапа: выявление концепций и когнитивные интервью [1, 55, 56]. В первую очередь с целью формирования совокупности элементов проводят когнитивные интервью или фокус-группы по выявлению концепций [51]. Руководства по структурированным интервью и обсуждениям фокус-групп обычно основываются на информации из обзора литературы и клинических интервью. На стадии выявления концепций пациенты способствуют обнаружению концепций и словаря, используемых для элементов процедуры PRO. В ходе выявления концепций для педиатрических процедур PRO осо-

бенно важно идентифицировать особый язык, с помощью которого пациенты описывают свою болезнь и лечение, поскольку дети и подростки могут пользоваться иными словами, нежели взрослые. Информацию, накопленную при выявлении концепций, следует использовать для поддержки концептуальной схемы инструмента, представляющей собой диаграмму, отражающую взаимоотношения между элементами, областями и концепциями, исследуемыми инструментом [1]. Важно, чтобы концептуальная схема педиатрического инструмента была специфична для целевой возрастной группы, а не являлась копией или аналогом концептуальной схемы для взрослых [14].

После составления схемы инструмента, вторым этапом является проведение когнитивных интервью с целью оценки и коррекции черновой версии инструмента на основании восприятия пациентами релевантности, чёткости и понятности материалов [52, 57]. Данные мероприятия по качественному анализу считают необходимыми для установления содержательной валидности, которую невозможно продемонстрировать путём количественного психометрического анализа. Руководство FDA даёт детальные рекомендации касательно методов определения содержательной валидности, но оно не рассматривает концептуальные или методологические проблемы, характерные для её оценки в педиатрических исследованиях.

Несмотря на призыв FDA к предоставлению документации по содержательной валидности, основанной на мнении пациента [1], исследователи по-прежнему избегают разрабатывать инструменты PRO на основе прямых данных от детей. Содержательная валидность педиатрических измерений, основанных на сообщениях детей и родителей, часто обосновывается данными, полученными не от детей, а от родителей или клинических экспертов [58—61]. Всё больше исследований, однако, указывают на то, что дети и подростки могут быть эффективными экспертами содержания инструмента [62—65]. Например, в одном исследовании изучали использование подростков как «опытных экспертов» при установлении содержательной валидности, и авторы пришли к выводу, что данные от подростков обеспечивали большую релевантность PRO для целевой популяции [66]. Понятно, что существуют ситуации, когда невозможно выяснить точку зрения ребёнка, к примеру, если целевая популяция слишком молода или если её когнитивный статус не позволяет читать, понимать или отвечать на вопросы о состоянии здоровья [67, 68]. Однако если дети способны обсуждать состояние своего здоровья, они могут обладать некой особенной точкой зрения, основанной на личном опыте, неизвестном клиницистам и родителям.

Как и для взрослых PRO, оценка точки зрения ребёнка в ходе выявления концепций и генерации элементов необходима для того, чтобы убедиться в релевантности и понятности материала педиатрических PRO для целевой аудитории. Более того, имеются некоторые аспекты восприятия и влияния симптомов, известные только самим пациентам [2]. Таким образом, мы рекомендуем либо включать детей в качественные исследования для установления валидности педиатрических PRO, либо предоставлять обоснование для исключения детей из исследования. При планировании и усовершенствовании материалов качественные данные от детей должны рассматриваться наряду с сообщениями родителей, клиницистов и/или экспертов.

Для педиатрических и взрослых PRO базовые принципы и методология валидности контента одинаковы. Оценка валидности контента для педиатрических материалов, тем не менее, связана с отдельными концептуальными сложностями и методологическими проблемами, не встречающимися при разработке материалов для взрослых популяций. Во-первых, ранние фазы установления валидности контента сфокусированы на выявлении целевой концепции материала PRO и «содержимого» элементов. При проведении исследования для выявления концепций и элементов для педиатрического инструмента необходимо рассмотреть центральную роль контекста в развитии ребёнка. Дети встроены во множество социальных сред, которые могут отличаться от таковых у взрослых. Эти среды включают семью, сверстников, школу, внешкольные занятия и общество [16, 69]. Влияние болезни и лечения, а также релевантность отдельных симптомов могут отличаться в каждой из этих сред. Например, некоторые симптомы относительно слабо влияют на взрослых, но в то же время ограничивают для ребёнка возможность участвовать в занятиях спортом или повседневными делами со сверстниками. Таким образом, при разработке педиатрических инструментов PRO, исследователи должны рассматривать релевантность и влияние симптомов и лечения внутри этих сред. Хотя родители могут иметь свою точку зрения на восприятие ребёнком своего здоровья дома, только сами дети могут дать полную картину, со всеми средами и социальными группами. Следовательно, чтобы обеспечить адекватность оценки инструментом всех важных аспектов заболевания и его лечения, при установлении содержательной валидности должна учитываться точка зрения ребёнка.

Центральная роль окружения ребёнка

Для педиатрических и взрослых PRO базовые принципы и методология валидности контента одинаковы. Оценка валидности контента для педиатрических материалов, тем не менее, связана с отдельными концептуальными сложностями и методологическими проблемами, не встречающимися при разработке материалов для взрослых популяций. Во-первых, ранние фазы установления валидности контента сфокусированы на выявлении целевой концепции материала PRO и «содержимого» элементов. При проведении исследования для выявления концепций и элементов для педиатрического инструмента необходимо рассмотреть центральную роль контекста в развитии ребёнка. Дети встроены во множество социальных сред, которые могут отличаться от таковых у взрослых. Эти среды включают семью, сверстников, школу, внешкольные занятия и общество [16, 69]. Влияние болезни и лечения, а также релевантность отдельных симптомов могут отличаться в каждой из этих сред. Например, некоторые симптомы относительно слабо влияют на взрослых, но в то же время ограничивают для ребёнка возможность участвовать в занятиях спортом или повседневными делами со сверстниками. Таким образом, при разработке педиатрических инструментов PRO, исследователи должны рассматривать релевантность и влияние симптомов и лечения внутри этих сред. Хотя родители могут иметь свою точку зрения на восприятие ребёнком своего здоровья дома, только сами дети могут дать полную картину, со всеми средами и социальными группами. Следовательно, чтобы обеспечить адекватность оценки инструментом всех важных аспектов заболевания и его лечения, при установлении содержательной валидности должна учитываться точка зрения ребёнка.

Выявление концепций и генерация элементов

Процедуры для выявления концепций и генерации элементов детально обсуждали ранее [51]. Однако исследования по выявлению концепций у детей сопряжены с дополнительными методологическими трудностями, не встречающимися при проведении

качественных исследований во взрослой популяции. Во-первых, мы рекомендуем проводить такие качественные исследования с разными типами респондентов, в том числе с самими детьми. Как указывалось, ранее, детей следует включать на как можно более ранних этапах установления содержательной валидности, особенно в ходе разработки инструмента фиксирования сообщений от детей. Ещё одним важным источником информации являются клиницисты, особенно педиатры, медсёстры педиатрических отделений и другие медицинские работники. При разработке педиатрических PRO необходимо проконсультироваться с клиницистами, имеющими опыт в лечении детей, а также в лечении данного заболевания. Родители или ответственные опекуны также являются существенным источником информации, особенно при разработке материалов для детей младшего возраста или анкет для родителей. Наконец, при оценке таких заболеваний как синдром дефицита внимания и гиперактивности, приобретающих особое значение в условиях школы, у учителей может быть особый взгляд на детские симптомы и их влияние на социальную и познавательную деятельность. В целом, содержательная валидность педиатрических PRO должна базироваться на качественных данных, полученных как от пациентов, так и от других информантов, например, родителей или клиницистов.

Исследования по выявлению концепций могут проводиться в формате индивидуальных интервью или фокус-групп, и каждый из этих подходов обладает преимуществами [51]. К примеру, фокус-группы позволяют исследователям опрашивать много пациентов за раз, в то время как пациенты могут пользоваться идеями других респондентов, чтобы высказывать свою точку зрения. У взрослых пациентов фокус-группы дают возможность участникам общаться друг с другом, что даёт больше информации о состоянии пациентов [56]. Однако некоторые проблемы, связанные с работой в группах, возникают чаще у детей и подростков, нежели у взрослых. Например, некоторые дети младшего возраста отвлекаются в большей степени, чем старшие дети или взрослые. Сообщения от детей могут быть подвержены влиянию сверстников или социально приемлемого смещения, что приводит к менее точному описанию симптомов и переживаний. Кроме того, из-за различного рода давления социума, возникающего в детстве и подростковом возрасте, педиатрические респонденты могут стесняться говорить откровенно в условиях группового обсуждения.

При проведении фокус-групп исследователи должны принимать меры, чтобы дети и подростки делились точной и правдивой информацией о своём здоровье. Например, хотя многие медицинские проблемы у взрослых могут обсуждаться в смешанных по полу группах, дети и подростки могут чув-

ствовать себя комфортнее и быть более открытыми в несмешанных группах, особенно при обсуждении таких тем как сексуальная активность или подростковые проблемы, связанные с взрослением. Исследователям также следует определить, может ли данное медицинское состояние вызывать стыд, неловкость или смущение. Например, подростки могут избегать обсуждения трудностей, которые они испытывают из-за ожирения, в кругу сверстников, и одно качественное исследование показало, что дети с заболеваниями стоп и голеностопных суставов стеснялись вида своей обуви, отличавшейся от обуви других детей [62]. Взрослые также могут стесняться обсуждать некоторые темы в присутствии других пациентов, но эта стеснительность или неловкость вызывает наибольшие сложности у детей и подростков.

Хотя индивидуальные интервью проигрывают методу фокус-групп по скорости, они также могут быть использованы для сбора более детальной информации о переживаниях индивидуума. Интервью могут быть более эффективны, чем фокус-группы, для чувствительных тем, вызывающих дискомфорт при обсуждении на публике. Кроме того, в отличие от результатов фокус-групп, данные индивидуальных интервью точно представляют каждого респондента при составлении таблицы или сетки насыщения для установления валидности контента (см. предыдущий отчёт исследовательской группы ISPOR о таблицах насыщения; *Patrick et al.* [51]). Напротив, при анализе данных фокус-групп, зачастую непросто выяснить точное число пациентов, сообщивших о некотором симптоме или воздействии.

Настоящая исследовательская группа рекомендует при разработке детских PRO проводить интервью по выявлению концепций у детей по возможности без родителей. Если во время интервью ребёнок чувствует себя некомфортно в отсутствие родителей, следует позволить родителям присоединиться к ребёнку в комнате интервью, но при этом не давать им участвовать в самом процессе беседы. Родителей также можно попросить сесть позади ребёнка, чтобы исключить возможность зрительного контакта с ребёнком или непреднамеренного влияния на его ответы. Кроме того, иногда опросник может быть составлен для пары родитель-ребёнок для оценки совместного опыта, связанного со здоровьем [70]. В таких ситуациях, когда исследуют совместную точку зрения, полезно разработать опросник, основанный на парных, а не на индивидуальных интервью по выявлению концепций. Также следует отметить, что имеются данные в пользу такого парного подхода в оценке HRQOL ребёнка, в отличие от фазы выявления концепций при разработке инструмента PRO. Некоторые инновационные исследования дают возможность предположить, что обученные интервьюеры могут получать надёжные и валидные ответы от пары родитель-ребёнок, при этом, возможно, повышая точность и сво-

дя к минимуму систематические ошибки и пробелы в ответах, что может происходить, когда дети самостоятельно проходят анкетирование или интервью [71, 72]. В то время как эти наводящие на размышления результаты заслуживают рассмотрения и дальнейшего изучения, настоящая исследовательская группа считает, что за редким исключением риски от участия родителей в интервью по выявлению концепций у детей перевешивают потенциальную выгоду. Следовательно, рекомендуют проводить интервью по выявлению концепций у детей по возможности без участия родителей (или участия других взрослых, например, учителей), чтобы быть уверенными в том, что ответы ребёнка не ошибочны и не ограничены мнением или присутствием родителя.

С учётом проблем, возникающих при проведении фокус-групп у детей и подростков, исследователи могут провести и фокус-группы, и индивидуальные интервью. Подобный комбинированный подход пользуется преимуществом больших объёмов информации, получаемых на основе групповой динамики в фокус-группах, подтверждая при этом, что индивидуальные интервью не дают новые концепции. Если фокус-группа оказывается непродуктивной для некоторых медицинских состояний в определённых возрастных группах, исследователям приходится продолжать качественные исследования с использованием исключительно индивидуальных интервью.

Часто разработка продуктивного интервью или фокус группы для выявления концепций у детей является трудной задачей. В начале интервью или фокус-групп интервьюер должен поговорить с детьми, чтобы убедиться, что они понимают цель беседы. Кроме того, прямые вопросы касательно симптомов и их влияния могут быть продуктивны только в том случае, если сформулированы и заданы в понятном для детей виде. Согласно одному из возможных подходов, начинать следует с более общих вопросов и постепенно переходить к более частным. Например, первые вопросы оказываются наиболее эффективными, если они ясны, непосредственны, просты и при этом касаются жизни ребёнка (например, «Расскажи, что ты делаешь дома после школы?»). Затем дополнительные вопросы могут оценивать влияние здоровья и болезни на занятия или же другие темы, затрагиваемые ребёнком в ответ на вводные вопросы. Последующие вопросы постепенно фокусируются конкретно на здоровье ребёнка и связанных с ним ограничениях. В ходе процесса выявления концепций, исследователи должны оставаться открытыми для изменений в структуре интервью, и на основе первых нескольких интервью может потребоваться внести изменения в руководства по их проведению.

Некоторые исследователи используют творчество, чтобы помочь детям дать полезную информацию при проведении интервью или фокус-групп. Интересным примером является исследование с фокус-груп-

пами, проведённое в ходе выявления концепций для Оксфордского Опросника о Стопе и Голеностопном Суставе (Oxford Ankle Foot Questionnaire), оценивающем нарушения, вызванные проблемами стопы и голеностопного сустава у детей [62]. В некоторых группах дети участвовали в «лайф-мэшинге», при этом их просили проанализировать один день из жизни гипотетического ребёнка с проблемами стопы или голеностопного сустава, в том числе трудности, возникающие в разное время дня и в разных ситуациях (например, утром, в школе, дома, на выходных и на каникулах). Детей просили дать имя этому выдуманному ребёнку и поработать вместе над описанием проблем, с которыми он сталкивается в каждом случае. Исследователи просили детей обсудить гипотетического ребёнка, чтобы впоследствии помочь им обсудить потенциально чувствительные темы, такие как боулинг, не спрашивая напрямую об их собственных сложных переживаниях. В качестве ещё одной стратегии, позволяющей детям чувствовать себя комфортно, эти фокус-группы начинали с разминки, включающей беседу на нейтральную тему. Такое решение положительно сказалось на желании детей свободно общаться в группе. Авторы пришли к выводу, что фокус-группы, включающие творческую деятельность детей, позволяли получать полезную информацию касательно последствий проблем стопы и голеностопного сустава в детстве даже у большинства детей самой младшей возрастной группы (5—7 лет).

В другом качественном исследовании использовали инновационный подход, включавший рисование, с целью получения описаний синдрома беспокойных ног и его последствий у детей в возрасте от 6 до 17 лет [73]. Детям задавали следующий вопрос: «Ты можешь нарисовать мне на листе бумаги, что ты ощущаешь, когда возникает это чувство в ногах?» После того, как рисунок закончен, к детям обращались со следующими словами: «Расскажи мне о своей картинке». По-видимому, рисунки способствовали налаживанию контакта и помогали детям описывать свои симптомы устно. Эти результаты дают основания предполагать, что рисование может быть потенциально полезным подходом в интервью или фокус-группах по выявлению концепций.

Когнитивные интервью

Стандартной практикой после планирования инструмента PRO является проведение интервью с респондентами из целевой аудитории, чтобы получить дополнительные сведения о содержательной валидности, подтвердив, что пациенты понимают элементы и верят в то, что опросник адекватно представляет интересующую концепцию [1, 52, 54]. Такие интервью часто называют «когнитивными итоговыми интервью», чаще — «когнитивными интервью». Для педиатрических инструментов мы рекомендуем проводить

когнитивные интервью с предполагаемым респондентом. Другими словами, дети должны быть проинтервьюированы касательно инструментов сообщений от детей, а родители — касательно родительских доверительных или наблюдательных инструментов.

В когнитивных интервью для взрослых инструментов респондентов из целевой популяции спрашивают о чёткости, вразумительности, полноте и релевантности элементов. Когнитивные интервью с детьми могут следовать тем же общим процедурам интервью, что были ранее описаны для взрослых [52, 55, 56, 74], но полезен также более тщательный распрос, чтобы убедиться, что интервьюер правильно понимает уровень понимания у данного ребёнка. Например, после того как ребёнок ответил на элемент, его могут попросить объяснить, как он или она выбрали такой ответ. Другими полезными вопросами об отдельных элементах могут быть «Как ты понимаешь этот вопрос?» и «Что тебя просят сделать в этом вопросе?». Ответы ребёнка помогут обнаружить, что ребёнок интерпретировал вопрос не так, как предполагал исследователь. Удобный подход для характеристики типов проблем, которые возникают у респондентов при работе с элементами, был предложен *Knafl и соавторами* [74].

Целесообразность разработки инструмента PRO — это основная задача, которую рассматривается когнитивными интервью. Руководство FDA [1] рекомендует «разработку инструментов в рамках довольно узких возрастных группировок». Эта рекомендация особенно важна для когнитивных интервью. Даже если инструмент PRO показал содержательную валидность в одном возрастном интервале, невозможно предполагать её наличие и в других интервалах. Хотя контент, необходимый для представления некоторой конструкции, может быть одинаковым для всех возрастных групп, возрастные различия в понимании отдельных элементов и в релевантности каждого из них могут влиять на содержательную валидность. Например, лексикон, широко используемый детьми одной возрастной группы для описания своих симптомов, может быть слишком сложным для младших детей. Более того, содержание некоторых элементов необязательно одинаково релевантно для всех детских возрастных групп, особенно элементов, оценивающих влияние симптомов на социальную сферу и занятия. Предыдущая исследовательская группа ISPOR дала детальные рекомендации по методам когнитивного интервьюирования для оценки других аспектов инструментов PRO, таких как конструкции, период воспроизведения, варианты ответа, форматирование инструмента, длительность инструмента и метод использования [52]. При проведении этих интервью для сообщаемых детьми инструментов, важно оценивать каждую из этих характеристик в рамках узких возрастных группировок, чтобы убедиться в том, что все аспекты опросника

PRO подходят для полного возрастного интервала предполагаемой популяции. В целом, и при разработке нового инструмента, и при использовании существующих инструментов для педиатрических выборок, когнитивные интервью должны документировать содержательную валидность, ясность, понятность и соответствие возрасту инструмента PRO в определённом возрастном интервале сообразно предполагаемой сфере применения инструмента.

В ходе когнитивного интервью дети могут испытывать большие по сравнению с взрослыми сложности в понимании своей роли как экспертов содержания в процессе разработки опросника. Например, некоторые дети могут фокусироваться на собственных ответах, а не на абстрактной задаче оценки содержания и ясности опросника [64]. Таким образом, в начале каждого интервью следует давать чёткое разъяснение хода оценки опросника, и интервьюеры должны уметь при необходимости концентрировать внимание ребёнка на задании по оценке.

Несмотря на небольшое число опубликованных исследований, проведённых с использованием когнитивного интервьюирования детей, некоторые из них дали обнадеживающие результаты. В ходе разработки педиатрического банка элементов, педиатрическая группа PROMIS провела когнитивные интервью с 77 детьми в возрасте 8-17 лет [47]. Дети во всех возрастных интервалах были в состоянии дать существенную информацию о своём здоровье и о самих элементах. Авторы также пришли к выводу, что дети также позволили составить представление о понятности элементов, так как дети младшего возраста помогли обнаружить несколько трудных для понимания элементов. Исследование с использованием когнитивных интервью PedsQL Diabetes Module также включало интервью с детьми 8 лет [75]. Версию опросника для детей 8-12 лет редактировали на основе вклада детей по результатам этих когнитивных интервью.

Надлежащая исследовательская практика 3: необходим ли инструмент для сообщаемых информантом исходов

При определении состояния здоровья младших детей важно рассматривать, помимо точки зрения ребёнка, ещё и мнение информантов. Анализ исходов, описываемых информантами, может понадобиться тогда, когда дети не способны сами правильно оформить показатели PRO по причине своего уровня развития, тяжести заболевания или когнитивного функционирования [16, 29, 76, 77]. Настоящая исследовательская группа использует термин «информант» по отношению к людям, не включая детей, которые предоставляют информацию о ребёнке. Чаще всего информантом является родитель ребёнка, но оценка также может быть проведена другими взрослыми, такими как клиницисты, учителя и

опекуны. Несмотря на то, что анкеты, заполненные информантами, фактически не являются «описанными пациентом», данная специальная группа предоставляет рекомендации по использованию данных от информантов из-за их частого применения в педиатрической оценке.

Прежде чем перейти к рекомендациям, необходимо определиться с соответствующей терминологией. Анкеты, заполненные информантами, могут быть разделены на две большие категории исходя из содержащихся в методе исследования заданий: доверительные или наблюдательные. Эти две категории могут быть далее разделены на подкатегории, как проиллюстрировано на рис. 3. Доверительные вопросы требуют от информантов выводов по детским субъективным впечатлениям. Многие формы, заполняемые родителями, включают доверительные задания, которые должны помочь выразить мнение ребенка через его субъективное восприятие, такое как эмоциональный статус, уровень удовлетворенности или степень болезненных ощущений [78—80]. Также информанта могут попросить описать, как ребенок отвечает на определенные вопросы. В одном исследовании, касавшемся опекунов взрослых пациентов, обнаружили, что вопросы, для ответа на которые информанту требуется представить себя на месте пациента, имеют большую схожесть с отчетами пациентов, чем вопросы к информантам о субъективных впечатлениях больных [81]. В педиатрическом исследовании доверительные вопросы могут принести полезные данные, так как родители, как правило, хорошо понимают своих детей, основываясь на жизненном опыте. Кроме того, мнение родителей имеет большое значение потому, что они обычно принимают решения в отношении детей и подростков.

Литература по доверительному анализу чаще основывается на степени совпадения между мнениями родителей и ребенка. Результаты исследований получились смешанными, причём некоторые испытания дали строгое соответствие [82—85], тогда как другие подчеркнули отличия во взглядах родителя и ребенка [86—92]. Степень соответствия определяется сложной комбинацией факторов, включающих здоровье родителя [86, 91, 93], тяжелое положение в семье [94], возраст ребенка [21, 86, 95], здоровье ребенка [25, 90, 93, 96], статистический метод для оценки соответствия [86, 93] и оцениваемую когнитивную сферу [25, 93]. С учётом этих данных, большинство исследователей признают, что отчеты родителей не могут считаться точно отражающими субъективное состояние здоровья ребенка. Несмотря на то, что родители могут давать полезную информацию о здоровье своих детей, для вопросов доверительного обзора требуются респондент, делающий заключение о субъективном мнении другого человека, и эти выводы не обязательно будут такими же, как истинное мнение обследуемого.

Вследствие этих трудностей, в руководстве FDA по PRO [1] не рекомендуют использование данных доверительной оценки исходов от респондентов для детского населения при изучении медицинской продукции. В руководство, однако, признается, что существуют педиатрические группы населения, такие как дети самого младшего возраста, которые не способны описать свое состояние здоровья. В ситуациях, когда необходимо получить информацию от информантов, FDA рекомендует, чтобы показатели были наблюдательными, а не доверительными. Руководство утверждает, что эти «наблюдательные описания... включают только те события или изменения поведения, которые могут быть отслежены». Для иллюстрации данной концепции FDA приводит следующий пример: «наблюдатели не могут достоверно рассказать об интенсивности болевых ощущений у младенца, но могут описать поведение младенца, предположительно изменившееся под влиянием болей». Определенные примеры наблюдаемых признаков детских симптомов приводят в руководстве FDA [97] по клинической разработке лекарств для лечения острого бактериального среднего отита, воспаления среднего уха, вызванного патогенными бактериями. Для младенцев и маленьких детей, которые не могут сформулировать свои симптомы, руководство рекомендует принимать во внимание рассказ взрослых о наблюдаемых признаках, таких как дергание уха, шелушение уха, возбужденность и сниженный аппетит.

Доверительные опросники приносят важную информацию о здоровье и развитии ребенка, и настоящая исследовательская группа призывает проводить исследования, включающие доверительные показатели от родителей. Тем не менее, исследовательская группа соглашается с рекомендациями FDA в том, что в исследованиях, проводимых для принятия управленческих решений, следует избегать использовать доверительные показатели. Данные, относящиеся к принятию таких решений и лицензированию медицинской продукции, должны быть по возможности непротиворечивыми. Результаты, основанные на доверительных опросниках, неизбежно содержат погрешности, как, например, противоречия между мнением информантов и детей.

Данная исследовательская группа также поддерживает рекомендации FDA по использованию наблюдательных показателей в необходимости применения отчетов информантов в исследованиях, связанных с лицензированием медицинской продукции. Несовпадение между ответами родителей и детей обычно в большей степени проявляется в недоступном наблюдении субъективных сферах, таких как боль и эмоциональные проблемы [9, 25, 98, 99]. Напротив, эти исследования зачастую обнаруживают строгое соответствие между ответами детей и родителей в измеряемых объективных областях, таких как соматическое функционирование. Данные выво-

ды подтверждают, что показатели отчетов родителей более точны и надежны, когда элементы опросника сосредоточены на наблюдательных материалах. Два примера наблюдательных анкет, разработанных для обследования маленьких детей, включают Индекс тошноты Келлера, который был создан для заполнения педиатрическими медицинскими сестрами [100], и заполняемый родителями пересмотренный Опросник по гастроэзофагеальному рефлюксу у младенцев [101]. Как видно на рис. 3, наблюдательные опросники могут быть касаться лишь текущей оценки поведения ребенка на момент наблюдения или при общем наблюдении за поведением ребенка в определенном временном промежутке. Любой из этих двух наблюдательных подходов подходит исследованиям, связанным с лицензированием медицинской продукции. Третий тип наблюдательных показателей, включающий записи высказываний ребенка, обычно менее предпочтителен, так как основывается на доверии во взаимоотношения между родителем и ребенком и на точности памяти родителей, относительно высказываний их детей.

Основываясь на анализе опубликованной литературы и недавних регуляторных изменений, настоящая исследовательская группа может дать три рекомендации относительно выбора между опросниками исходов, заполненными родителями либо детьми в управленческой сфере. Во-первых, когда дети в целевом

возрастном диапазоне способны сами воспользоваться инструментом по PRO, тогда показатели, полученные от детей, следует использовать (для обсуждения по изменяющимся различиям в способности выполнения показателей PRO был изложен в Надлежащей исследовательской практике 1). В регламентирующей среде предпочтительны опросники, заполненные ребенком, потому что это наиболее точная оценка переживания ребенком заболевания и лечения без какой-либо погрешности или интерпретации, которая может исказить данные при использовании отчетов информантов. Чтобы быть способными пользоваться инструментом по PRO, дети в целевом возрастном диапазоне должны быть в состоянии прочитать опросник, понять основные положения и ответить на них адекватно и достоверно. Чтение и понимание должны быть проверены с помощью когнитивного опроса, во время которого адекватность и правильность ответов может быть зафиксирована в больших психометрических исследованиях правильности показателей PRO. Из-за различий в развитии внутри любой возрастной группы, всегда есть несколько детей, которые не могут выполнить оценку PRO, в отличие большинства их сверстников. Поэтому исследователи должны определять крайние значения и придерживаться основных тенденций внутри целевой возрастной группы при выяснении, способна ли возрастная группа использовать отдельный инструмент по PRO.



Рис. 3. Виды опросников, заполняемых информантами, для оценки в педиатрической практике

Во-вторых, если дети в целевом возрастном диапазоне не могут выполнить анализ PRO, следует использовать показатели, описанные информантами. В-третьих, когда используется опросник, заполненный информантом, все элементы инструмента должны по возможности соответствовать наблюдению, а не субъективному восприятию ребенка.

Также необходимо отметить, что есть ситуации, когда могут понадобиться и родительские, и детские формы одного и того же инструмента, заполняемые одновременно. Интерпретация данных из двух одновременных источников, однако, зачастую поднимает вопросы, подобные обсуждаемым в разделе «Выводы и рекомендации для будущих исследований» настоящего обзора.

Надлежащая исследовательская практика 4: убедиться, что инструмент создан и оформлен соответственно целевой возрастной группе

В некоторых обзорных статьях обсуждаются вопросы дизайна и оформления параметров педиатрических PRO, различающихся по возрасту респондентов [6, 8, 9, 14–16, 24, 102]. В ответ на уклон в сторону качественных методов в руководстве FDA по PRO, возрастает число исследований, непосредственно принимающих во внимание пожелания детей по структуре и оформлению опросника [19, 47, 103, 104]. Такие предварительные исследования привнесли полезную информацию, и наше понимание, каким должны быть дизайн и формат опросника, будет увеличиваться по мере публикации все большего количества педиатрических когнитивных интервьюированных исследований. Кроме того, мы поддерживаем проведение более качественных испытаний, направленных на поиск наиболее эффективных дизайна и оформления параметров PRO.

В последующих разделах обсуждается перечень факторов, который должен быть учтен, чтобы гарантировать, что инструмент по PRO соответствует рассматриваемой педиатрической возрастной группе. Важно принять во внимание эти факторы при разработке нового, а также при применении ранее проверенного инструмента в новой возрастной группе. Определенные возрастные ограничения для любого аспекта дизайна исследования основываются на множестве взаимосвязанных факторов, таких как предметный состав и связанные со здоровьем характеристики целевой группы населения. Поэтому данный раздел не рассматривает возрастные ограничения, применяемые в каждой ситуации. Вместо этого, целью главы является обзор аспектов дизайна и оформления инструмента, которые с большей вероятностью будут различаться среди возрастных групп при выполнении рекомендаций, основанных на современных тенденциях, описанных в опубликованных исследованиях.

Связанный со здоровьем словарный запас и уровень чтения

Зачастую отмечается, что у детей словарный запас, связанный со здоровьем, увеличивается с возрастом. Сведения, опубликованные Rebok и соавторами, иллюстрируют данную тенденцию. В данном исследовании понимание детьми терминов, относящихся к здоровью, было проверено внутри четырех возрастных групп: возраста 5, 6, 7 и с 8 до 11 лет (рис. 2). Для большинства терминов процент респондентов с малым пониманием снижался по мере увеличения возраста. Например, 14,3% пятилетних детей слабо осознают значение термина «боль», в сравнении с 8,3% шестилетних, 6,2% семилетних и 0% старших детей. Процент недостаточного знания других понятий, связанных со здоровьем, также снижается с возрастом, например, «энергия», «достаточное здоровье», «состояние комфорта» и «нервный». Похожие примеры выводов были описаны в ряде других педиатрических исследований [71].

Ясно, что словарный запас и читабельность опросника не может быть выше лингвистических способностей целевой популяции. Разработчики инструмента долгое время избегали этой проблемы, но делали попытки упростить язык и понятия в версиях, созданных специально для младших возрастных групп [9, 80, 105]. Известный метод заключается в оценке читабельности опросника, используя показатели читабельности, такие как оценка уровня качества Flesch-Kinkaid, и подгонке опросника к целевому возрастному диапазону [9, 80, 106, 107]. В дополнение к терминологии, факторы, относящиеся к читабельности, включают: сложность синтаксиса, длину предложений и число предложений в каждом параграфе [14]. Одно направление разработки опросника с соответствующим возрасту словарным составом состоит в соблюдении языка, используемого детьми во время опроса и в целевых группах [51]. Тогда словарный запас, полученный напрямую от детей из целевого возрастного диапазона, может быть включен в окончательный инструмент.

При подаче педиатрических показателей PRO, вместе с заявлением на регистрацию, важно иметь документацию о пригодности уровня языка для использования в целевой возрастной группе. Соответствие возраста уровню читаемости и словарному запасу может быть подкреплено баллами индекса читабельности и данными когнитивных опросов, в которых респонденты целевой возрастной группы предоставляют обратную связь по доступности вопросов для понимания. Методология когнитивных опросов была признана эффективным способом для коррекции словаря банка педиатрических вопросов PROMIS [47]. В таком качественном интервьюированном исследовании 77 детей и подростков в возрасте от 8 до 17 лет, когнитивные опросы выявляли

некоторые термины, которые были непонятны для части детей в данной возрастной группе. Сложные термины включали понятия: раздражительный (рекомендуемая замена: капризный), изможденный (рекомендуемая замена: уставший), взбешенный (рекомендуемая замена: злой или свирепый), общественные мероприятия (рекомендуемая замена: отдых с друзьями). Похожие результаты были получены в исследовании, включавшем когнитивные опросы, направленные на диабетический модуль PedsQL [75]. Дети сообщили, что некоторые слова были слишком сложными. По этой причине версия опросника для 8- и 12-летних детей была скорректирована для включения языка, соответствующего возрасту. Например, «утомление» было изменено на «усталость», а «раздражительный» было изменено на «капризный или угрюмый».

Шкала ответной реакции

Обзор связанных со здоровьем описанных пациентами показателей для детей в возрасте от 3 до 8 лет обнаружил, что шкалы Лайкерта (то есть шкалы с характеристиками по всему диапазону) чаще всего использовались в качестве шкал ответной реакции в 34 из 53 опросников из данной возрастной группы [24]. Другие типы шкал включают графическое отображение, выражение лица и наглядную аналоговую шкалу, которая очень редко используется в педиатрических критериях и зачастую непонятна для маленьких детей [110]. Одно исследование 120 детей в возрасте от 6 до 18 лет выявило, что дети предпочитают шкалы Лайкерта наглядным аналоговым шкалам и пронумерованным 10-балльным шкалам, объясняя это тем, что шкалы Лайкерта проще для заполнения [104]. Более того, когнитивное интервьюированное исследование, изучавшее банк педиатрических вопросов PROMIS, выявило, что даже самые маленькие дети в группе (например, 8-лет) не имели видимых трудностей в понимании вариантов выбора в шкале Лайкерта из пяти пунктов [47]. В дополнение, комиссия Delphi, созданная для разработки детских опросников HRQOL, пришла к единому заключению с учетом мнения 94,1% респондентов, что шкалы Лайкерта наиболее подходят для применения у детей [111].

Резюмируя, шкалы Лайкерта адекватны большинству педиатрических опросников PRO, что было подтверждено качественными исследованиями, мнением экспертов, мнением детей и значительным числом опросников, использовавших шкалы в прошлом. Однако элементы со шкалами Лайкерта не подойдут любому педиатрическому опроснику для любой возрастной группы. В двух исследованиях были получены достоверные выводы в отношении возрастных различий в способности детей использовать все параметры шкалы Лайкерта. Маленькие дети склонны

предоставлять больше информации на самом высоком и самом низком уровне шкалы ответной реакции. В одном из таких исследований дети от 5 до 6 лет предоставляли более радикальные ответы, чем более взрослые дети при оценке эмоционально окрашенных заданий по трех- и пятибалльной шкале [44]. Другое исследование обнаружило, что процент крайних ответов снижается с возрастом: 87,1% в возрасте 5 лет, 78,9% в возрасте 6 лет, 61,4% в возрасте 7 лет и 50,4% в возрасте от 8 до 11 лет [19]. Данные этих двух испытаний показывают, что в некоторых случаях шкалы ответной реакции Лайкерта могут быть уместными для старших, но не младших детей. Поэтому мы рекомендуем оценивать варианты ответов по критериям PRO с помощью детей целевой возрастной группы в когнитивных интервью. В них важно установить возраст, в котором дети начинают использовать и понимать весь спектр параметров шкалы ответной реакции, а не только крайние ответы. Чтобы подходить для опросников PRO, используемых в заявлениях на регистрацию, параметры ответов должны также отвечать критериям, описанным в руководстве FDA по PRO [1], включающим отличия в выборе ответов, ясные формулировки и понятные инструкции (возможно включающие пример заполненного ответа на вопрос).

Период воспроизведения

Недавний обзор инструментов оценки HRQOL у детей показал, что многие часто используемые опросники, заполняемые детьми, имеют период воспроизведения в течение последних 4 недель или последнего месяца [14]. К ним относят CHIP (Опросник по здоровью детей, Child Health Questionnaire) [79, 112, 15] и PedsQL [113]. Опросники, такие как эти три, с относительно длительными периодами воспроизведения привносят важную информацию о здоровье детей во многих клинических испытаниях и других исследованиях. Однако инструменты с периодами воспроизведения не более месяца могут быть скептически оценены при использовании в заявке на регистрацию продукта.

В руководстве FDA по PRO [1] говорится:

Инструменты PRO, которые зависят от памяти пациентов, в особенности, если они должны предоставлять информацию за длительный период,... вероятнее всего будут снижать содержательную валидность. Ответная реакция с большой вероятностью будет зависеть от состояния пациента во время воспроизведения. По этим причинам, вопросы с короткими периодами воспроизведения или вопросы, требующие от пациентов описать их настоящее или недавнее состояние, являются предпочтительными.

В руководстве FDA по PRO не приводят оптимальный период воспроизведения, потому что он зависит от большого количества факторов, включа-

ющих предназначение инструмента, а также продолжительность, частоту и интенсивность оцениваемой концепции. Руководство отмечает, что адекватность периода следует оценивать с учетом «способности пациента к правильному воспроизведению требуемой информации».

Есть некоторая особенность в том, что способность детей к правильному использованию периодов воспроизведения различается среди возрастных групп. Например, одно исследование обнаружило, что дети младше 8 лет, по сравнению со старшими, были менее способны рассказать о последнем появлении симптома в последние 4 недели, и было не ясно, правильно ли 5- и 6-летние дети понимают значение недели или месяца [19]. Исследование пар родитель-ребенок также подтвердило ограничения в использовании периодов воспроизведения у детей [71]. В этом испытании у детей иногда проявлялся недостаток когнитивных навыков, необходимых для оценки временных рамок, и родители часто помогали своим детям вспоминать случаи проблем со здоровьем, связывая временной промежуток с определенными событиями и занятиями. Результаты были распределены не по возрасту, но 69% исследуемых были от 8 до 10 лет, предполагая, что некоторые дети в этом диапазоне могли иметь трудности с периодами воспроизведения. Другое недавнее исследование проверяло эффект «сохранения в памяти» (то есть прошедшее время между проведением обследования и описываемым событием) по точности памяти детей на приемы пищи [114]. В данном примере на детях четвертого поколения (то есть примерно 9 лет) прочность памяти на школьные завтраки и ланчи была улучшена путем укорочения времени запоминания. Это открытие подчеркивает важность использования более коротких периодов воспроизведения для признаков, описываемых детьми. Разработчики инструмента PRO опасаются, что дети могут иметь больше трудностей с периодами воспроизведения, чем взрослые. Например, Опросник по качеству жизни детей с астмой (Child Asthma Quality of Life Questionnaire) имеет период воспроизведения длительностью в 1 неделю, который был уменьшен с 2 недель во взрослой версии инструмента [115].

Резюмируя, более короткие периоды воспроизведения являются предпочтительными для оценки PRO, используемой в управленческой сфере, и они могут быть важнее для педиатрических опросников, чем для взрослых. Данная исследовательская группа считает, что периоды воспроизведения длительностью 24 часа или менее выглядят предпочтительнее, чем более длительные в регламентирующей сфере, основанной на рекомендациях руководства FDA по PRO по более коротким периодам воспроизведения и времени исследования, выявляющих сложности у детей в понимании продолжительных временных интервалов. Более того, способ моментальной оцен-

ки (то есть оценка представлений ребенка о своём текущем состоянии) может быть полезен, так как он помогает избежать многих проблем, возникающих при непонимании детьми периодизации времени.

Невозможно, однако, предложить такой определенный период воспроизведения, который был бы оптимальным для всех педиатрических исследований, при условии, что длительные периоды могут быть применимы лишь в адекватно обоснованных ситуациях. Короткие периоды воспроизведения могут также иметь недостатки, такие как необходимость частого анализа и возможное упущение при анализе важных, имеющих отношение к здоровью симптомов или событий, происходящих за пределами определенных периодов. Также, наличие периодов воспроизведения, согласованных с клиническими параметрами в отдельном испытании, может быть выгодно для сбора данных и их интерпретации для оценки PRO. Некоторые терапевтические области, такие как психиатрия, стремятся использовать опросники, заполняемые клиницистами, с периодом воспроизведения в 1 неделю, а PRO с более коротким периодом воспроизведения могут негативно влиять на интерпретацию полученных сведений. Все эти факторы должны быть рассмотрены при определении адекватного периода воспроизведения для каждой задачи, инструмента и возрастной группы. Для информирования будущих решений о периодах воспроизведения необходимо исследование детской памяти с использованием различных временных рамок, которое следует сочетать с группировкой по узкому диапазону возрастов.

Продолжительность применения инструмента

Объем педиатрического опросника PRO требует тщательного обсуждения из-за больших различий в способности детей концентрировать внимание на заданиях [16]. Обычно считается, что инструменты, разработанные для использования у маленьких детей, имеют большее количество ограничений, чем разработанные для использования у старших детей и взрослых. Также предполагается, что для маленьких детей требуются более короткие инструкции, чем для старших [47]. Излишне длительные испытания могут стать причиной упущения детьми данных или несерьезного отношения к каждому пункту, тем самым, предоставляя менее точные и надежные сведения [9]. При рассмотрении этих проблем, одним способом решения было создание множества версий одного инструмента, различающихся по продолжительности и нацеленных на разные возрастные группы [105]. Пробное тестирование и когнитивные опросы детей в целевой возрастной группе могут помочь определить, соответствует ли целевой возрастной группе продолжительность использования инструмента перед включением его в клиническое испытание.

Наглядное представление

Разработчики инструмента часто включают наглядное изображение методов и возможных вариантов действий в анализе педиатрических PRO [14]. Небольшое количество эмпирических данных демонстрирует преимущества этих иллюстраций, но в ходе одного из исследований обнаружили, что педиатрический опросник с иллюстрациями значительно быстрее заполнялся, чем вариант без иллюстраций, что наводит на мысль, что мультипликационная версия гораздо эффективнее привлекает внимание детей [116]. Несмотря на небольшое эмпирическое обоснование, многие исследователи считают, что данный наглядный подход способен помочь маленьким детям пройти опросник. Теоретически, картинки могут помочь пробудить интерес у детей, удержать их внимание, улучшить доступность вопросов для понимания, прояснить, как именно нужно отвечать, и, следовательно, способствовать более целенаправленным ответам [9]. Для проверки и подтверждения таких потенциальных преимуществ необходимы новые исследования, но, основываясь на мнениях экспертов и многих инструментах, использующих данный подход, становится ясно, что картинки могут быть полезными для маленьких детей.

Применяют несколько графических подходов. Простейшим способом для конкретизации вариантов ответа может быть использование окружностей возрастающих размеров. Например, в CHIP-CE от детей требуется выбрать один из предложенных вариантов для ответа на вопросы [112]. Каждому ответу соответствует определенная окружность, которые увеличиваются в размере по мере перехода вариантов от «никогда» до «всегда» (рис. 4). Возможные ответы могут также быть изображены в виде рисунков выражений лиц, которые отображают нейтральность или различные степени позитивных или негативных эмоций (рис. 5) [105, 117—119]. Такая

методика «улыбающегося лица» рекомендована для использования у маленьких детей 80% участниками комиссии, управляющей разработкой показателей качества жизни KIDSCREEN [111].

Поэтому, регламентирующие органы со скептицизмом смотрят на данную схему. Воспроизведено из Med Care, 42, Riley AW, Forrest CB, Rebok GW, et al. The Child Report Form of the CHIP-Child Edition: reliability and validity, 221-31, 2004, with permission from Lippincott Williams & Wilkins. Существуют также

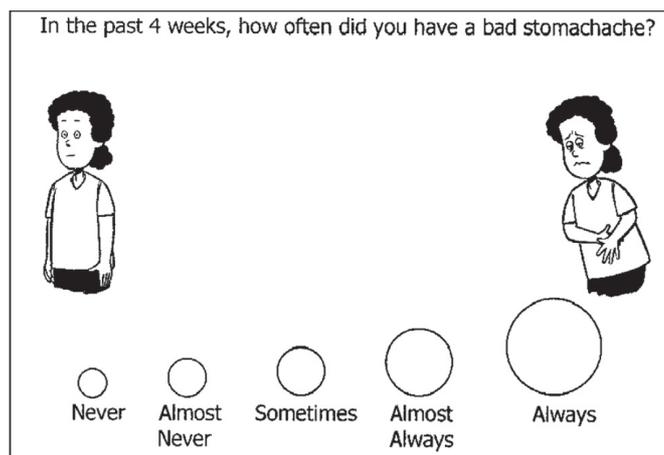


Рис. 4. Пример шкалы ответов, использующей иллюстрации и окружности увеличивающихся размеров для облегчения выбора варианта ответа

*В последние 4 недели, как часто у вас появлялась боль в животе? Ответ: никогда, почти никогда, временами, почти постоянно, постоянно Следующий примером вопроса по PRO с использованием наглядного представления, помогающего детям понять, как работать со шкалой. Однако, другие аспекты этой схемы не оптимальны для применения в регламентирующей сфере. Например, относительно долгий период воспроизведения и параметры ответов требуют от ребёнка усреднить его впечатления о симптомах за 4-недельный промежуток времени. Это сложная задача, а память детей не настолько точная для охвата 4-недельного периода.



Рис. 5. Пример шкалы ответов с выражениями лиц, отражающими варианты ответа

*Какое из этих лиц наиболее точно описывает ваши ощущения во рту или горле в данный момент? Обведите одно. Ответы: нет боли (0), чуть-чуть болит (1), болит немного сильнее (2), болит ещё сильнее (3), сильно болит (4), очень сильно болит (5).

Воспроизведено из Eur J Oncol Nurs, 14, Tomlinson D., Gibson F., Treister N., et al., Refinement of the Children's International Mucositis Evaluation Scale (ChIMES): child and parent perspectives on understandability/, content validity and acceptability, 41, 2010, with permission from Elsevier.

примеры опросников, использующих более детализированные рисунки для иллюстрации вопросов и вариантов ответа. CHIP-CE использует изображения детей, отражающих крайние варианты по шкале ответов Лайкерта для каждого вопроса (рис. 4) [112]. Другие критерии используют мультипликационные иллюстрации, предназначенные, чтобы помочь детям понять содержание вопроса, включающие изображения детей, отражающих психиатрические диагнозы (рис. 6) [120], и картинки с собаками, показывающие педиатрические дерматологические симптомы [116].

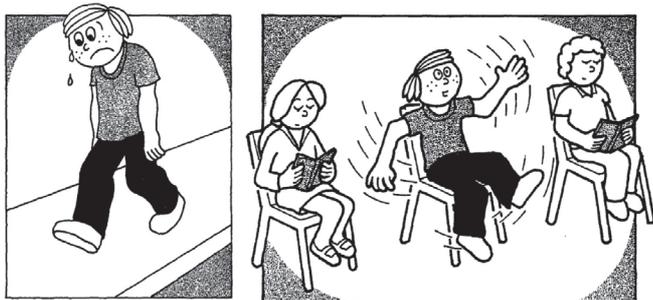


Рис. 6. Примеры рисунков, используемых для иллюстрации психиатрических заболеваний в заполняемом ребёнком опроснике

Воспроизведены из J. Abnorm Child Psychol, 22, Valla J.P., Bergeron L., Berube H, et al., A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6-11 years): development, validity, and reliability, 403-23, 1994. With kind permission from Springer Science and Business Media.

Результаты одного из исследований, однако, показывают, что испытуемым следует использовать эти графические методы с осторожностью. Chambers and Craig [118] сравнили два вида шкал, классифицирующих боль, используя лица в качестве вариантов ответов в группе из 100 детей. Одна из шкал включала нейтральное лицо «нет боли» как точку отсчета, в то время как другая в качестве этой же точки отсчета использовала улыбающееся лицо (рис. 7). Два типа шкал привели к получению значительно различающихся сумм баллов по вопросам. Эти выводы пока-

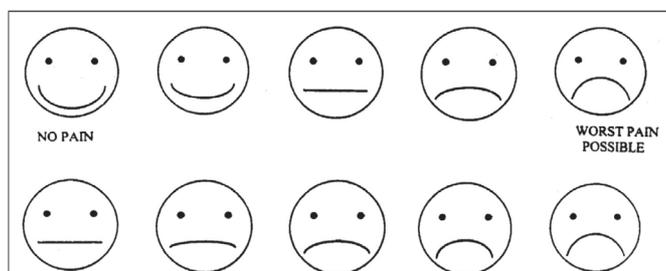


Рис. 7. Два типа шкал, использующих лица, с различными выражениями для понятия «нет боли»

Ответы: нет боли, наибольшая боль.
Воспроизведены из Pain, 78, Chambers CT, Craig KD, An intrusive impact of anchors in children's faces pain scales, 27-37, 1998, with permission from the International Association for the Study of Pain.

зывают, что иллюстрации должны быть разработаны и проверены также тщательно, как и другие составные части опросника, так как изображения могут исказить результаты и снизить их ценность. Другим возможным риском использования иллюстраций является то, что некоторые подростки могут посчитать их слишком детскими. Так, когнитивные интервью могут быть использованы для проверки восприятия и понимания детьми наглядных изображений вместе с другими аспектами структуры опросника.

Другие детали оформления

Другие детали оформления, которые на первый взгляд могут показаться менее важными, также имеют влияние на наглядность и ценность показателей педиатрических PRO. Выработка четкого плана вопросов с достаточной величиной шрифта может улучшить возможность прочтения, в частности, для маленьких детей. Вопросы по формату редко обсуждаются в публикуемой литературе, но в некоторых публикациях указывают, что в данном вопросе не выработалось общепринятое мнение. Например, по результатам качественных интервью было пересмотрено оформление Анкеты по влиянию здоровья полости рта у детей (Child Oral Health Impact Profile), и был увеличен размер букв и затенения в каждом вопросе [103]. Более того, в исследовании с когнитивными интервью для банка вопросов PROMIS дети-респонденты предложили сделать буквы больше и выделить жирным шрифтом период воспроизведения для большей ясности [47].

Организационные подходы

При разработке и внедрении показателей, описанных взрослыми пациентами, исследователи обычно исходят из того, что большинство больных может заполнить опросник самостоятельно. Однако дети различаются по своей способности самостоятельно заполнить опросник. Вследствие этого, при разработке показателей педиатрических PRO, исследователям необходимо принимать во внимание все виды управленческих подходов, различающихся по степени самостоятельности, ожидаемой от ребенка-респондента. Предполагается, что старшие дети и подростки, в большинстве случаев могут сами заполнить опросник. Для маленьких детей могут потребоваться критерии, заполняемые опрашиваемым, или помощь родителей в преодолении ограничений в способности к чтению и концентрации внимания. Например, Опросник по детской астме был разработан в трех видах, соответствующих трем возрастным периодам [105, 119]. Форма для самого младшего возраста (4-7 лет) заполняется ребенком с помощью родителей, форма для среднего возрастного диапазона (8-11 лет) может быть заполнена при необходимости при участии взрослых, и форма для

старшей возрастной группы (12-16 лет) была создана для самостоятельного выполнения (смотри «Выводы и рекомендации для будущих исследований» ниже в обзоре, где обсуждаются подходы, комбинирующие данные из множества версий инструмента, определенных для каждого возраста). Другие виды управленческих подходов, таких как электронный сбор данных (смотри ниже) или использование дополнительных средств может повысить внимание детей и осознание ими процесса ответа на вопросы [24].

Выбор управленческого подхода зависит от многих факторов, включающих возраст детей и сложность оцениваемых концепций. Следует подробно оговорить административный подход во время качественной фазы разработки инструмента. Например, если по результатам когнитивных опросов предполагается, что детям сложно самим выполнить задание, возможно, что менее самостоятельный организационный подход принесет больше полезной информации. Но, как уже обсуждалось в Надлежащей исследовательской практике 2, присутствие родителей или другого взрослого может исказить или препятствовать ответам ребенка. Поэтому настоящая исследовательская группа рекомендует, что родителям или другим взрослым следует помогать детям при заполнении инструмента PRO, только когда это действительно необходимо.

Для обеспечения успешного сбора данных необходимо, чтобы члены коллектива, управляющие опросами и другими параметрами исследования, были достаточно квалифицированы. Обучение особенно важно для неограниченных испытаний, таких как опросы для оценки подходов, требующие от интервьюера большей чувствительности и понимания. Процесс опроса ребенка и подростка может существенно отличаться от опроса взрослого. Интервьюеры должны иметь базовые знания по разработанным вопросам, а также способам взаимоотношений с детьми и их поведенческим реакциям [121]. Рекомендации по опросу родителей и учителей о поведении детей и его объяснению также были опубликованы ранее [121].

Несмотря на то, что исследовательская группа анализировала, главным образом, организацию PRO, стоит отметить, что исследования в педиатрии требуют повышенного внимания ко всем аспектам процесса сбора данных. Например, исследователям необходимо разработать соответствующий возрасту язык для объяснения детям целей испытания и порядка его проведения. Перед началом выполнения детьми оценки PRO, интервьюеры или другие участники исследования должны информировать их о том, что от них потребуется, сообщать цель задаваемых вопросов, предполагаемое использование данных, конфиденциальность процесса и что делать при появлении неприятных ощущений или как прекратить свое участие. Пока дети не понимают деталей исследования в том же объеме, что и взрослые, исследователям следу-

ет пытаться дать детям хотя бы общее представление о нём и позволить им чувствовать себя комфортно.

Электронный сбор данных

Электронные методики сбора данных PRO (то есть эPRO или ePRO) могут иметь значительные преимущества в исследованиях с детьми и подростками. Устройства с дисплеями, полезные при работе с детьми, включают портативные устройства, такие как смартфоны, планшеты и сенсорные экраны, работающие без использования клавиатуры или мыши, стационарные или переносные компьютеры [122]. В некоторых опубликованных исследованиях говорится о том, что прохождение детьми электронных опросников, включающих управление через Интернет, может быть целенаправленным, надежным и эффективным способом сбора данных [123—127]. Многие дети хорошо знакомы и уверенно себя чувствуют при работе с экраном, и в одном исследовании было обнаружено, что дети предпочитают проходить опросы по собственному здоровью и поведению через Интернет, а не использовать бумажные версии [123]. Если способы заполнения при помощи экрана могут помочь детям оставаться сконцентрированными и увлеченными, то они будут способствовать улучшению качества данных, полученных непосредственно от пациента, при уменьшении количества потерянной информации. Технология интерактивного голосового ответа может также стать подходящим способом сбора данных для части возрастных групп [128]. Этот удобный метод, включающий словесную озвучку вопросов с помощью телефона, может быть проще для выполнения, чем напечатанные опросники для детей с ограниченными способностями к чтению и письму.

Предыдущие специальные группы ISPOR по PRO продвигали рекомендации по разработке и тестированию показателей эPRO, включая когнитивные опросы, тестирование на практичность и оценку идентичности бумажных и электронных версий опросника [52, 122]. Эти рекомендации по разработке эPRO имеют отношение и к опросникам у детей и подростков. Кроме того, настоящая специальная группа рекомендует оценивать пригодность к использованию методов электронного сбора данных в пределах небольших возрастных групп при создании показателей PRO для детей. Как и в случае напечатанного опросника, заполняемого самостоятельно пациентом, технология, основанная на использовании дисплеев, опирается на способности респондента к чтению, но применение эPRO может быть эффективным в одной возрастной группе, а не в нескольких. Вследствие этого, исследователям необходимо документировать сферы применения показателей ePRO по всем соответствующим возрастным группам, чтобы они были признаны эффективными в регламентирующих органах.

Надлежащая исследовательская практика 5: обсуждение межкультурных аспектов

При создании педиатрических инструментов для широкого использования важно определить возможные различия, связанные с культурными и языковыми особенностями. Несмотря на то, что методы для перевода и подтверждения эффективности педиатрических показателей в различных культурных слоях общества сходны с используемыми во взрослой практике [129—131], могут возникать дополнительные вопросы, связанные с их применением. Например, из-за различий в системе образования разных стран способность детей к чтению будет отличаться в любом рассматриваемом возрасте. Отличия в текстах на разных языках могут также приводить к географическому разнообразию уровня грамотности внутри определенных возрастных групп [132]. Поэтому может потребоваться повторная оценка соответствия инструмента возрасту в каждой новой стране, где он может применяться у детей. Также культурные различия могут быть в виде информации, сообщаемой детям о заболевании и его лечении, а также отличия по степени самостоятельности детей в описании своего состояния. Кроме того, могут иметься различия в готовности детей общаться с интервьюерами без родителей. Подобные культурные нормы требуют новых исследований по определению содержательной валидности в пределах каждой новой культуры. Наконец, влияние заболевания и лечения на социальное, эмоциональное и качественное функционирование может сильно отличаться в разных культурах, завися от обыденной деятельности детей данной культуры.

Вследствие этого, целесообразность инструмента по педиатрическим PRO требует пересмотра, вместе с оценкой валидности материалов для каждой нации. Оценка должна проводиться с учетом всех значимых особенностей инструмента, включая инструкции, цели, подходы, словарный состав и наглядные изображения. В целом, разработка межкультурного инструмента PRO для детей, вероятно, потребует большего внимания и труда, чем простое следование межкультурным руководствам, написанным для инструментов для взрослых. Будущие исследования и руководства должны помочь в определении шагов, которые нужно пройти при переводе и оценке опросников PRO в разных культурах.

Выводы и рекомендации для будущих исследований

Предоставляемые исследовательской группой рекомендации не следует рассматривать в качестве жестких правил, обязательных к исполнению в каждом исследовании. На самом деле, целью исследовательской группы было создание общего руковод-

ства и обсуждение важных вопросов, касающихся разработки, оценки или внедрения инструментов педиатрических PRO. Полученную информацию предполагают использовать при подаче заявок на регистрацию и при составлении инструкции по медицинскому применению медицинской продукции. Каждое отдельное исследование по-своему уникально, и оптимальный способ оценки зависит от множества факторов, включающих возраст ребенка, состояние здоровья исследуемой группы населения и вид оцениваемой модели. Более того, различные подходы к PRO и дизайну исследования могут быть рекомендованы к изучению, не направленному на поддержку регистрации медицинской продукции и решений регулирующих органов. Вследствие этого, надлежащие исследовательские практики, рекомендуемые в настоящем обзоре, следует рассматривать как точку отсчета, и каждый исследователь должен учитывать цель и условия проведения каждого испытания.

Исследование PRO у детей и подростков является относительно новым направлением изучения, и необходима дальнейшая работа по подготовке усовершенствованных инструментов PRO и методических руководств для будущих исследований. Например, необходимы исследования по оптимизации дизайна PRO у маленьких детей, в особенности у детей младше 8 лет, для которых самостоятельно описанные показатели характеризуются неустойчивой надежностью и валидностью. Существуют определенные пути формулировки, представления и оформления вопросов для улучшения их доступности для понимания у маленьких детей и, соответственно, увеличения надежности и валидности инструмента. Исследования, сравнивающие различные подходы к оценке, могут способствовать разработке более целенаправленных рекомендаций, чем доступные на настоящий момент. Также, больше усилий необходимо для изучения качественных проектов исследования, которые с большей вероятностью будут эффективны у маленьких детей. В данном обзоре предлагаются основные рекомендации для качественных методов, но лишь небольшое количество исследований, проверявших пригодность инструмента по PRO для детей, было опубликовано. Поэтому нужны дополнительные исследования для проверки и улучшения этих методик. Эта методология развивается по мере проведения и публикации все большего количества качественных исследований в данной группе населения.

При использовании инструментов описанных информантом исходов все большее внимание уделяется разработке истинно обсервационных вопросов, вместо репрезентативных, требующих выводов о субъективном мнении ребенка. Следовательно, для отражения данных обсервационных подходов может быть полезным улучшение и обоснование применения часто используемых инструментов сообщений от пациентов или клиницистов.

Другая проблема заключается в интерпретации данных от различных возрастных групп. Многие опросники PRO у детей создаются в виде множества версий для разных возрастных групп. Содержание вопросов, шкалы ответов, словарный состав, наглядное изображение вариантов ответа и другие характеристики инструмента могут отличаться среди различных вариантов для определенных возрастов, обеспечивая соответствие каждого варианта своей возрастной группе. Респонденты также могут быть разными: родители, отвечающие за маленьких детей, и старшие дети, заполняющие опросник самостоятельно. Такой множественный подход может гарантировать, что определенный инструмент будет отражать соответствующее нужному возрасту содержание с порядком оценки, соответствующей широкому возрастному диапазону. Целью исследования PRO, при его проведении в соответствии с задачами регулирующих органов, является создание выборочных инструкций по применению, а не индивидуальных для каждой узкой возрастной группы. Как следствие, использование разных вариантов ставит вопрос о том, как данные из различных вариантов или даже различных докладчиков могут быть использованы при написании одной инструкции для детей всех возрастов. Одним возможным аналитическим подходом является суммирование данных из множества источников в одном анализе, но для этого необходимо обеспечить, чтобы содержание и параметры оценки всех версий были эквивалентны и сравнимы. Другой подход — исследовать разные версии отдельно друг от друга и проверить, сравнимы ли закономерности исходов лечения, в целом по разным возрастам и вариантам инструмента. Нужны дальнейшие исследования для разработки руководств по проведению анализа и интерпретации данных, собранных с использованием множества версий одного инструмента по PRO, специфичных для конкретного возраста.

Также может быть сложно интерпретировать данные от родителей и детей, заполняющих одинаковые формы одного и того же инструмента. В некоторых ситуациях и родители, и дети могут давать полезную информацию, но не до конца ясно, как должны рассматриваться результаты, если два респондента имеют разные взгляды на одни и те же инструменты. Более того, в связи с увеличивающимся вниманием к ограниченным опросникам, заполняемым информантами, как к значимой сфере (смотри Надлежащую исследовательскую практику 3), некоторые формы сообщений от пациентов могут иметь несколько иное содержание по сравнению с формами, заполняемыми детьми и включающими больше субъективной информации. Поэтому при получении данных от родителей и детей лучше всего будет определить априори, что данные от одного из докладчиков принимаются за точку отсчета. Необходимо

больше исследований для создания руководства, охватывающего мнения и родителей, и детей по оценке исходов, даже при их сильном различии.

Новые исследования также необходимы для изучения чувствительности инструментов по педиатрическим PRO. Чувствительность — это диапазон, в котором опросники состояния здоровья точно отражают изменения в здоровье пациента на протяжении времени [133—135]. С точки зрения регламентирующих решений чувствительность необходима для соответствия опросника PRO цели «в определении различий в баллах в течение времени у отдельных лиц и групп, у которых есть изменения с отношением к оцениваемому подходу» [1]. Как отмечалось в обсуждении Надлежащей исследовательской практики 1 в настоящем обзоре, надежность и валидность была проанализирована и описана для многих показателей PRO, разработанных для детей и подростков. Однако опубликовано относительно немного данных по чувствительности показателей педиатрических PRO [14]. С учётом важности точного обнаружения изменений в исследованиях, связанных с маркировкой медицинской продукции, настоящая исследовательская группа рекомендует оценивать чувствительность любого педиатрического опросника PRO, используемого в данном контексте.

Несмотря на то, что данный обзор специальной группы ориентируется, в первую очередь, на создание новых инструментов PRO, рекомендации и пояснения могут также относиться к использованию, оценке и модификации существующих опросников PRO, разработанных для применения у детей. Вопросы, касающиеся использования и совершенствования существующих инструментов, детально рассматривались предыдущей исследовательской группой ISPOR [54].

Одним из недостатков данной специальной группы было ограниченное внимание к оценке педиатрических PRO в сфере исследований, связанных с подачей документации в регулирующие органы. Исследования исходов здоровья вне данной области не включали в настоящий обзор, но стоит отметить, что другие области исследования у детей и подростков также поднимают важные методологические вопросы. Например, оценка практичности структур здравоохранения необходима для определения количества лет жизни с учетом ее качества (QALY) при моделировании «затраты-полезность», согласованным со страховыми организациями и другими лицами, принимающими решения. Но оценка состояния здоровья детей поднимает особые вопросы [136—138]. Дети могут быть не способны представить себе абстрактные понятия, необходимые для непосредственной оценки свойств, таких как временные согласованные и стандартные игровые задания, и зачастую методы оценки свойств, используемые у взрослых, не применимы к педиатрическим

заболеваниям. Более того, в связи с ускоренным физическим и умственным развитием детей, состояние детского здоровья менее стабильно, чем взрослого, а временная оценка здоровья сложна [139]. Тем не менее, анализ «затраты-полезность» зачастую связан со схемами лечения детской патологии, и средства для педиатрических состояний необходимы для оценки качества жизни детей с учётом данных схем [140]. Поэтому исследователи работают над созданием практических опросников здоровья, специфичных для детей и подростков [141—144]. Другое расширяющееся поле для исследований, которое в будущем будет влиять на клинические и политические решения — это исследование сравнительной эффективности (Comparative Effectiveness Research, CER). Предполагается, что методы и понятия CER могут отличаться при оценке лечения у детей, по сравнению с взрослыми [145]. По причине определенного уклона в сторону PRO в регламентирующей сфере, настоящая исследовательская группа не рассматривает данные области исследований, но ожидается, что будущие испытания продолжат изучать вопросы, касающиеся детей, по широкому ряду тем, таких как практическая оценка состояния здоровья и CER.

Многие другие важные темы, относящиеся к педиатрическим исследованиям, не были освещены в настоящем обзоре. Например, оценка лекарственной терапии у детей затрагивает специфические этические вопросы, детально рассмотренные в опубликованных руководствах Американской педиатрической академии (American Academy of Pediatrics) [146]. Вопросы, касающиеся привлечения детей для рандомизированного клинического исследования, также были рассмотрены ранее [147].

Литература

1. Food and Drug Administration. Guidance for Industry — Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Silver Spring, MD: Food and Drug Administration, 2009.
2. Patrick D.L., Burke L.B., Powers J.H., et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007;10 (Suppl. 2):S125—37.
3. Burke L.B., Kennedy D.L., Miskala P.H., et al. The use of patient-reported outcome measures in the evaluation of medical products for regulatory approval. *Clin Pharmacol Ther* 2008;84:281—3.
4. Gnanasakthy A., Mordin M., Clark M., et al. A review of patient-reported outcome labels in the United States: 2006 to 2010. *Value Health* 2012;15:437—42.
5. European Medicines Agency. Reflection Paper on the Regulatory Guidance for the Use of Health-Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products. London, United Kingdom: European Medicines Agency. Committee for Medicinal Products for Human Use (CHMP), 2005.
6. Connolly M.A., Johnson J.A. Measuring quality of life in paediatric patients. *Pharmacoeconomics* 1999;16:605—25.
7. Drotar D. Measuring Health-Related Quality of Life in Children and Adolescents: Implication for Research and Practice. Mahwah, NJ: Erlbaum, 1998.
8. Eiser C., Mohay H., Morse R. The measurement of quality of life in young children. *Child Care Health Dev* 2000;26:401—14.
9. Eiser C., Morse R. Quality-of-life measures in chronic diseases of childhood. *Health Technol Assess* 2001;5:1—157.
10. Eiser C., Morse R. A review of measures of quality of life for children with chronic illness. *Arch Dis Child* 2001;84:205—11.
11. Erling A. Methodological considerations in the assessment of health-related quality of life in children. *Acta Paediatr Suppl* 1999;88:106—7.
12. De Civita M., Regier D., Alamgir A.H., et al. Evaluating health-related quality-of-life studies in paediatric populations: some conceptual, methodological and developmental considerations and recent applications. *Pharmacoeconomics* 2005;23:659—85.
13. Hinds P.S., Brandon J., Allen C., et al. Patient-reported outcomes in end-of-life research in pediatric oncology. *J Pediatr Psychol*, 32 (2007), pp. 1079—1088
14. Landgraf J.M. Practical considerations in the measurement of HRQL in child/adolescent clinical trials. P. Fayers, R. Hays (Eds.), *Assessing Quality of Life in Clinical Trials* (2nd ed.), Oxford University Press, New York (2005)
15. Landgraf J.M., Abetz L.N. Measuring health outcomes in pediatric populations: issues in psychometrics and application. *Qual Life Pharmacoecon Clin Trials*, 2 (1996), pp. 793—802
16. Matza L.S., Swensen A.R., Flood E.M., et al. Assessment of health-related quality of life in children: a review of conceptual, methodological, and regulatory issues. *Value Health*, 7 (2004), pp. 79—92
17. Morris C., Gibbons E., Fitzpatrick R. Child and Parent Reported Outcome Measures: A Scoping Report Focusing on Feasibility for Routine Use in the NHS. Department of Public Health, University of Oxford, Oxford, United Kingdom (2009).

С каждым новым исследованием PRO у детей и подростков уверенность в инструментах и методологии по педиатрическим PRO будет продолжать расти. Более того, по мере роста использования данных по PRO для написания инструкций, все больше информации будет появляться о стандартах применения инструментов. В связи с непрекращающимся развитием данных областей исследования, использование инструментов PRO в клинических исследованиях и заявки на регистрацию медицинского продукта помогут убедиться в том, что переживание ребенком болезни и процесса лечения адекватным образом отображены и проанализированы при принятии управленческих решений.

Благодарности

Авторы искренне благодарны следующим членам ISPOR за их вклад в разработку этих совместных надлежащих исследовательских практик, указанных в данном отчете: *Elizabeth Molsen, Bhagwan Aggarwal, Celestina Arrigo, Alexandra Barsdorf, Katy Benjamin, Andrew Bottomley, Yan Cheng, Hsingwen Chung, Ben Craig, Dalia Dawoud, Prasanna Deshpande, Pero Draganic?, Ancilla Fernandes, Emmuela Flood, Mary Gawlicki, Richa Goyal, Rosina Hinojosa, Stefan Holmstrom, Manthan Janodia, Peter Kaskel, David Keleti, Kanav Khera, Ramanath Kv, Jeanne Landgraf, Kathryn Lasch, Tamar Lasky, Yvonne Lis, Michelle Mocariski, Chris Morris, Shekoufeh Nikfar, Mira Pavlovic, Farhan Abdul Rauf, Yong Joo Rhee, Margaret Rothman, Joerg Ruof, Steve Ryder, Luciana Scalone, Jun Su, Tara Symonds, Elio Tanaka, Sreedhar Tirunagari, Nayanabhirama Udupa, Wendy Ungar, Erica Velthuis, and Etta Vinik.*

18. *Quittner A.L., Modi A., Cruz I.* Systematic review of health-related quality of life measures for children with respiratory conditions. *Paediatr Respir Rev*, 9 (2008), pp. 220—232.
19. *Rebok G., Riley A., Forrest C., et al.* Elementary school-aged children's reports of their health: a cognitive interviewing study. *Qual Life Res*, 10 (2001), pp. 59—70.
20. *Theunissen N.C., Vogels T.G., Koopman H.M., et al.* The proxy problem: child report versus parent report in health-related quality of life research. *Qual Life Res*, 7 (1998), pp. 387—397.
21. *Varni J.W., Limbers C.A., Burwinkle T.M.* How young can children reliably and validly self-report their health-related quality of life? An analysis of 8,591 children across age subgroups with the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*, 5 (2007), p. 1.
22. *Wallander J.L., Schmitt M., Koot H.M.* Quality of life measurement in children and adolescents: issues, instruments, and applications. *J Clin Psychol*, 57 (2001), pp. 571—585.
23. *Clarke S.A., Eiser C.* The measurement of health-related quality of life (QOL) in paediatric clinical trials: a systematic review. *Health Qual Life Outcomes*, 2 (2004), p. 66.
24. *Cremeens J., Eiser C., Blades M.* Characteristics of health-related self-report measures for children aged three to eight years: a review of the literature. *Qual Life Res*, 15 (2006), pp. 739—754.
25. *Eiser C., Morse R.* Can parents rate their child's health-related quality of life? Results of a systematic review. *Qual Life Res*, 10 (2001), pp. 347—357.
26. *Rajmil L., Herdman M., Fernandez de Sanmamed M.J., et al.* Generic health-related quality of life instruments in children and adolescents: a qualitative analysis of content. *J Adolesc Health*, 34 (2004), pp. 37—45.
27. *Ravens-Sieberer U., Erhart M., Wille N., et al.* Generic health-related quality-of-life assessment in children and adolescents: methodological considerations. *Pharmacoeconomics*, 24 (2006), pp. 1199—1220.
28. *Solans M., Pane S., Estrada M.D., et al.* Health-related quality of life measurement in children and adolescents: a systematic review of generic and disease-specific instruments. *Value Health*, 11 (2008), pp. 742—764.
29. *Bevans K.B., Riley A.W., Moon J., et al.* Conceptual and methodological advances in child-reported outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res*, 10 (2010), pp. 385—396.
30. *Riley A.W.* Evidence that school-age children can self-report on their health. *Ambul Pediatr*, 4 (2004), pp. 371—376.
31. *Williams C.* Gender, adolescence and the management of diabetes. *J Adv Nurs*, 30 (1999), pp. 1160—1166.
32. *Patrick D.L., Edwards T.C., Skalicky A.M., et al.* Validation of a quality-of-life measure for deaf or hard of hearing youth. *Otolaryngol Head Neck Surg*, 145 (2011), pp. 137—145.
33. *Peden M.* World report on child injury prevention appeals to "Keep Kids Safe". *Inj Prev*, 14 (2008), pp. 413—414.
34. Office of the United Nations High Commissioner for Human Rights. *Convention on the Rights of the Child*. Geneva, Switzerland: United Nations, 1989.
35. Centers for Disease Control and Prevention. *Protect the Ones You Love: Child Injuries Are Preventable*. Atlanta, GA: Centers for Disease Control and Prevention, 2012.
36. *Borse N.N., Rudd R.A., Dellinger A.M., et al.* Years of Potential Life Lost Due to Unintentional Injuries among Children and Adolescents. Atlanta, GA: Centers for Disease Control and Prevention, 2012.
37. US Department of Health & Human Services. *What Is Adolescence?* Washington, D.C.: US Department of Health & Human Services, Office of Population Affairs, 2012.
38. *Kaplan P.S.* Adolescence. Houghton Mifflin Company, Boston (2004).
39. *Grange A., Bekker H., Noyes J., et al.* Adequacy of health-related quality of life measures in children under 5 years old: systematic review. *J Adv Nurs*, 59 (2007), pp. 197—220.
40. *Khan F.I., Reddy R.C., Baptist A.P.* Pediatric Dyspnea Scale for use in hospitalized patients with asthma. *J Allergy Clin Immunol*, 123 (2009), pp. 660—664.
41. *Luby J.L., Belden A., Sullivan J., et al.* Preschoolers' contribution to their diagnosis of depression and anxiety: uses and limitations of young child self-report of symptoms. *Child Psychiatry Hum Dev*, 38 (2007), pp. 321—338.
42. *Riley A., Forrest C.B., Starfield B., et al.* Technical Manual for the Child Health and Illness Profile-Child Edition and CHIP-CE. The Johns Hopkins University, Baltimore, MD (2001)
43. *Ronen G.M., Streiner D.L., Rosenbaum P., et al.* Health-related quality of life in children with epilepsy: development and validation of self-report and parent proxy measures. *Epilepsia*, 44 (2003), pp. 598—612.
44. *Chambers C.T., Johnston C.* Developmental differences in children's use of rating scales. *J Pediatr Psychol*, 27 (2002), pp. 27—36.
45. *Demetriou A., Raftopoulos A.* Modeling the developing mind: from structure to change. *Dev Rev*, 19 (1999), pp. 319—368.
46. *Juniper E.F., Guyatt G.H., Feeny D.H., et al.* Minimum skills required by children to complete health-related quality of life instruments for asthma: comparison of measurement properties. *Eur Respir J*, 10 (1997), pp. 2285—2294.
47. *Irwin D.E., Varni J.W., Yeatts K., et al.* Cognitive interviewing methodology in the development of a pediatric item bank: a Patient Reported Outcomes Measurement Information System (PROMIS) study. *Health Qual Life Outcomes*, 7 (2009), p. 3.
48. *Edwards T.C., Huebner C.E., Connell F.A., et al.* Adolescent quality of life, part I: conceptual and measurement model. *J Adolesc*, 25 (2002), pp. 275—286.
49. *Arnett J.J.* Emerging adulthood: a theory of development from the late teens through the twenties. *Am Psychol*, 55 (2000), pp. 469—480.
50. *Henig R.M.* What is it about 20-somethings? *The New York Times*. August 22 (2010) MM28.
51. *Patrick D.L., Burke L.B., Gwaltney C.J., et al.* Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 1—eliciting concepts for a new PRO instrument. *Value Health*, 14 (2011), pp. 967—977.
52. *Patrick D.L., Burke L.B., Gwaltney C.J., et al.* Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: part 2—assessing respondent understanding. *Value Health*, 14 (2011), pp. 978—988.
53. *Fehnel S.* Establishing optimal requirements for content validity: a work in progress. *Value Health*, 12 (2009), p. 1074
54. *Rothman M., Burke L., Erickson P., et al.* Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. *Value Health*, 12 (2009), pp. 1075—1083.
55. *Brod M., Tesler L.E., Christensen T.L.* Qualitative research and content validity: developing best practices based on science and experience. *Qual Life Res*, 18 (2009), pp. 1263—1278.
56. *Leidy N.K., Vernon M.* Perspectives on patient-reported outcomes: content validity and qualitative research in a changing clinical trial environment. *Pharmacoeconomics*, 26 (2008), pp. 363—370.
57. *Collins D.* Pretesting survey instruments: an overview of cognitive methods. *Qual Life Res*, 12 (2003), pp. 229—238.
58. *Deane D.M.* Content, construct, and criterion-related validity. *Diabetes Educ*, 17 (1991), pp. 361—362.
59. *Gannotti M.E., Cruz C.* Content and construct validity of a Spanish translation of the Pediatric Evaluation of Disability Inventory for children living in Puerto Rico. *Phys Occup Ther Pediatr*, 20 (2001), pp. 7—24.
60. *McCabe M.A., Granger C.V.* Content validity of a pediatric functional independence measure. *Appl Nurs Res*, 3 (1990), pp. 120—122.
61. *Penny A.M., Waschbusch D.A., Klein R.M., et al.* Developing a measure of sluggish cognitive tempo for children: content validity, factor structure, and reliability. *Psychol Assess*, 21 (2009), pp. 380—389.

62. Morris C., Liabo K., Wright P., et al. Development of the Oxford Ankle Foot Questionnaire: finding out how children are affected by foot and ankle problems. *Child Care Health Dev*, 33 (2007), pp. 559—568.
63. Riesch S.K., Anderson L.S., Angresano N., et al. Evaluating content validity and test-retest reliability of the children's health risk behavior scale. *Public Health Nurs*, 23 (2006), pp. 366—372.
64. Stewart J.L., Lynn M.R., Mishel M.H. Evaluating content validity for children's self-report instruments using children as content experts. *Nurs Res*, 54 (2005), pp. 414—418.
65. Tomlinson D., Gibson F., Treister N., et al. Understandability, content validity, and overall acceptability of the Children's International Mucositis Evaluation Scale (ChIMES): child and parent reporting. *J Pediatr Hematol Oncol*, 31 (2009), pp. 416—423.
66. Schilling L.S., Dixon J.K., Knafl K.A., et al. Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. *Nurs Res*, 56 (2007), pp. 361—366.
67. Harris S.R., Daniels L.E. Content validity of the Harris Infant Neuromotor Test. *Phys Ther*, 76 (1996), pp. 727—737.
68. Palisano R.J., Rosenbaum P., Bartlett D., et al. Content validity of the expanded and revised Gross Motor Function Classification System. *Dev Med Child Neurol*, 50 (2008), pp. 744—750.
69. Cox M.J., Paley B. Families as systems. *Annu Rev Psychol*, 48 (1997), pp. 243—267.
70. Pleil A.M., Kimel M., McCormack J., et al. Psychometric Assessment of the Injection Pen Assessment Questionnaire (IPAQ): measuring ease of use and preference with injection pens for human growth hormone. *Health Qual Life Outcomes*, 10 (2012), p. 126.
71. Ungar W.J., Mirabelli C., Cousins M., et al. A qualitative analysis of a dyad approach to health-related quality of life measurement in children with asthma. *Soc Sci Med*, 63 (2006), pp. 2354—2366.
72. Ungar W.J., Boydell K., Dell S., et al. A parent-child dyad approach to the assessment of health status and health-related quality of life in children with asthma. *Pharmacoeconomics*, 30 (2012), pp. 697—712.
73. Picchetti D.L., Arbuckle R.A., Abetz L., et al. Pediatric restless legs syndrome: analysis of symptom descriptions and drawings. *J Child Neurol*, 26 (2011), pp. 1365—1376.
74. Knafl K., Deatrick J., Gallo A., et al. The analysis and interpretation of cognitive interviews for instrument development. *Res Nurs Health*, 30 (2007), pp. 224—234.
75. Varni J.W., Curtis B.H., Abetz L.N., et al. Content validity of the PedsQL 3.2 Diabetes Module in newly diagnosed patients with type 1 diabetes mellitus ages 8-45 [published online ahead of print December 27, 2012]. *Qual Life Res* (2012) 10.1007/s11136-012-0339-8
76. Chang P.C., Yeh C.H. Agreement between child self-report and parent proxy-report to evaluate quality of life in children with cancer. *Psychooncology*, 14 (2005), pp. 125—134.
77. Varni J.W., Limbers C.A., Burwinkle T.M. Parent proxy-report of their children's health-related quality of life: an analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL 4.0 Generic Core Scales. *Health Qual Life Outcomes*, 5 (2007), p. 2.
78. Landgraf J., Abetz L., Ware J. The CHQ User's Manual. (2nd ed.) Health Act, Boston (1999).
79. Riley A.W., Forrest C.B., Starfield B., et al. The Parent Report Form of the CHIP-Child Edition: reliability and validity. *Med Care*, 42 (2004), pp. 210—220.
80. Varni J.W., Seid M., Kurtin P.S. Pediatric health-related quality of life measurement technology: a guide for health care decision makers. *JCOM*, 6 (1999), pp. 33—40.
81. Pickard A.S., Lin H.W., Knight S.J., et al. Proxy assessment of health-related quality of life in African American and white respondents with prostate cancer: perspective matters. *Med Care*, 47 (2009), pp. 176—183.
82. Beattie P.E., Lewis-Jones M.S. A comparative study of impairment of quality of life in children with skin disease and children with other chronic childhood diseases. *Br J Dermatol*, 155 (2006), pp. 145—151.
83. Brown G.T., Wright F.V., Lang B.A., et al. Clinical responsiveness of self-report functional assessment measures for children with juvenile idiopathic arthritis undergoing intraarticular corticosteroid injections. *Arthritis Rheum*, 53 (2005), pp. 897—904.
84. Brunner H.I., Klein-Gitelman M.S., Miller M.J., et al. Health of children with chronic arthritis: relationship of different measures and the quality of parent proxy reporting. *Arthritis Rheum*, 51 (2004), pp. 763—773.
85. Hutchings H.A., Upton P., Cheung W.Y., et al. Development of a parent version of the Manchester-Minneapolis quality of life survey for use by parents and carers of UK children: MMQL-UK (PF). *Health Qual Life Outcomes*, 6 (2008), p. 19.
86. Creemeens J., Eiser C., Blades M. Factors influencing agreement between child self-report and parent proxy-reports on the Pediatric Quality of Life Inventory 4.0 (PedsQL) generic core scales. *Health Qual Life Outcomes*, 4 (2006), p. 58.
87. Davis E., Nicolas C., Waters E., et al. Parent-proxy and child self-reported health-related quality of life: using qualitative methods to explain the discordance. *Qual Life Res*, 16 (2007), pp. 863—871.
88. De Los Reyes A., Kazdin A.E. Measuring informant discrepancies in clinical child research. *Psychol Assess*, 16 (2004), pp. 330—334.
89. Jozefiak T., Larsson B., Wichstrom L., et al. Quality of life as reported by school children and their parents: a cross-sectional survey. *Health Qual Life Outcomes*, 6 (2008), p. 34.
90. Kiss E., Kapornai K., Baji I., et al. Assessing quality of life: mother-child agreement in depressed and non-depressed Hungarian. *Eur Child Adolesc Psychiatry*, 18 (2009), pp. 265—273.
91. Vance Y.H., Morse R.C., Jenney M.E., et al. Issues in measuring quality of life in childhood cancer: measures, proxies, and parental mental health. *J Child Psychol Psychiatry*, 42 (2001), pp. 661—667.
92. Zhou H., Roberts P., Horgan L. Association between self-report pain ratings of child and parent, child and nurse and parent and nurse dyads: meta-analysis. *J Adv Nurs*, 63 (2008), pp. 334—342.
93. Upton P., Lawford J., Eiser C. Parent-child agreement across child health-related quality of life instruments: a review of the literature. *Qual Life Res*, 17 (2008), pp. 895—913.
94. Canning E.H., Hanser S.B., Shade K.A., et al. Maternal distress and discrepancy in reports of psychopathology in chronically ill children. *Psychosomatics*, 34 (1993), pp. 506—511.
95. Guyatt G.H., Juniper E.F., Griffith L.E., et al. Children and adult perceptions of childhood asthma. *Pediatrics*, 99 (1997), pp. 165—168.
96. Levi R.B., Drotar D. Health-related quality of life in childhood cancer: discrepancy in parent-child reports. *Int J Cancer Suppl*, 12 (1999), pp. 58—64.
97. Food and Drug Administration. Guidance for Industry—Acute Bacterial Otitis Media: Developing Drugs for Treatment. Food and Drug Administration, Silver Spring, MD (2012).
98. Janse A.J., Sinnema G., Uiterwaal C.S., et al. Quality of life in chronic illness: children, parents and paediatricians have different, but stable perceptions. *Acta Paediatr*, 97 (2008), pp. 1118—1124.
99. Klassen A.F., Miller A., Fine S. Agreement between parent and child report of quality of life in children with attention-deficit/hyperactivity disorder. *Child Care Health Dev*, 32 (2006), pp. 397—406.
100. Keller V.E., Keck J.F. An instrument for observational assessment of nausea in young children. *Pediatr Nurs*, 32 (2006), pp. 420—426.
101. Kleinman L., Rothman M., Strauss R., et al. The infant gastroesophageal reflux questionnaire revised: development and validation as an evaluative instrument. *Clin Gastroenterol Hepatol*, 4 (2006), pp. 588—596.
102. Annett R.D. Assessment of health status and quality of life outcomes for children with asthma. *J Allergy Clin Immunol*, 107 (Suppl.) (2001), pp. S473—S481.
103. Broder H.L., McGrath C., Cisneros G.J. Questionnaire development: face validity and item impact testing of the Child Oral Health Impact Profile. *Community Dent Oral Epidemiol*, 35 (Suppl. 1) (2007), pp. 8—19.

104. van Laerhoven H., van der Zaag-Loonen H.J., Derckx B.H. A comparison of Likert scale and visual analogue scales as response options in children's questionnaires. *Acta Paediatr*, 93 (2004), pp. 830—835.
105. Christie M.J., French D., Sowden A., et al. Development of child-centered disease-specific questionnaires for living with asthma. *Psychosom Med*, 55 (1993), pp. 541—548.
106. Flesch R. A new readability yardstick. *J Appl Psychol*, 32 (1948), pp. 221—233.
107. Prout H.T., Chizik R. Readability of child and adolescent self-report measures. *J Consult Clin Psychol*, 56 (1988), pp. 152—154.
108. Likert R.A. A technique for the development of attitude scales. *Educ Psychol Meas*, 12 (1952), pp. 313—315.
109. Streiner D.L., Norman G.R. *Health Measurement Scales: A Practical Guide to Their Development and Use*. Oxford University Press, New York (2008).
110. Shields B.J., Palermo T.M., Powers J.D., et al. Predictors of a child's ability to use a visual analogue scale. *Child Care Health Dev*, 29 (2003), pp. 281—290.
111. Herdman M., Rajmil L., Ravens-Sieberer U., et al. Expert consensus in the development of a European health-related quality of life measure for children and adolescents: a Delphi study. *Acta Paediatr*, 91 (2002), pp. 1385—1390.
112. Riley A.W., Forrest C.B., Rebok G.W., et al. The Child Report Form of the CHIP-Child Edition: reliability and validity. *Med Care*, 42 (2004), pp. 221—231.
113. Varni J.W., Seid M., Kurtin P.S. PedsQL 4.0: reliability and validity of the Pediatric Quality of Life Inventory version 4.0 generic core scales in healthy and patient populations. *Med Care*, 39 (2001), pp. 800—812.
114. Baxter S.D., Guinn C.H., Royer J.A., et al. Accuracy of children's school-breakfast reports and school-lunch reports (in 24-h dietary recalls) differs by retention interval. *Eur J Clin Nutr*, 63 (2009), pp. 1394—1403.
115. Juniper E.F., Guyatt G.H., Feeny D.H., et al. Measuring quality of life in the parents of children with asthma. *Qual Life Res*, 5 (1996), pp. 27—34.
116. Holme S.A., Man I., Sharpe J.L., et al. The Children's Dermatology Life Quality Index: validation of the cartoon version. *Br J Dermatol*, 148 (2003), pp. 285—290.
117. Tomlinson D., Gibson F., Treister N., et al. Refinement of the Children's International Mucositis Evaluation Scale (ChIMES): child and parent perspectives on understandability, content validity and acceptability. *Eur J Oncol Nurs*, 14 (2010), pp. 29—41.
118. Chambers C.T., Craig K.D. An intrusive impact of anchors in children's faces pain scales. *Pain*, 78 (1998), pp. 27—37.
119. French D.J., Christie M.J., Sowden A.J. The reproducibility of the Childhood Asthma Questionnaires: measures of quality of life for children with asthma aged 4-16 years. *Qual Life Res*, 3 (1994), pp. 215—224.
120. Valla J.P., Bergeron L., Berube H., et al. A structured pictorial questionnaire to assess DSM-III-R-based diagnoses in children (6-11 years): development, validity, and reliability. *J Abnorm Child Psychol*, 22 (1994), pp. 403—423.
121. Whitcomb S.A., Merrell K.W. *Interviewing techniques*. Behavioral, Social, and Emotional Assessment of Children and Adolescents (4th ed.) Routledge, New York (2013) chapter 6.
122. Coons S.J., Gwaltney C.J., Hays R.D., et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. *Value Health*, 12 (2009), pp. 419—429.
123. Mangunkusumo R.T., Duisterhout J.S., de Graaff N., et al. Internet versus paper mode of health and health behavior questionnaires in elementary schools: asthma and fruit as examples. *J Sch Health*, 76 (2006), pp. 80—86.
124. Mussaffi H., Omer R., Prais D., et al. Computerised paediatric asthma quality of life questionnaires in routine care. *Arch Dis Child*, 92 (2007), pp. 678—682.
125. Raat H., Mangunkusumo R.T., Landgraf J.M., et al. Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Qual Life Res*, 16 (2007), pp. 675—685.
126. Varni J.W., Limbers C.A., Burwinkle T.M., et al. The ePedsQL in type 1 and type 2 diabetes: feasibility, reliability, and validity of the Pediatric Quality of Life Inventory Internet administration. *Diabetes Care*, 31 (2008), pp. 672—677.
127. Young N.L., Varni J.W., Snider L., et al. The Internet is valid and reliable for child-report: an example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). *J Clin Epidemiol*, 62 (2009), pp. 314—320.
128. Stritzke W.G., Dandy J., Durkin K., et al. Use of interactive voice response (IVR) technology in health research with children. *Behav Res Methods*, 37 (2005), pp. 119—126.
129. Bullinger M., Alonso J., Apolone G., et al. Translating health status questionnaires and evaluating their quality: the IQOLA Project approach. *International Quality of Life Assessment*. *J Clin Epidemiol*, 51 (1998), pp. 913—923.
130. Wild D., Grove A., Martin M., et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. *Value Health*, 8 (2005), pp. 94—104.
131. Guillemin F., Bombardier C., Beaton D. Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *J Clin Epidemiol*, 46 (1993), pp. 1417—1432.
132. Smythe I., Everatt J., Al-Menaye N., et al. Predictors of word-level literacy amongst grade 3 children in five diverse languages. *Dyslexia*, 14 (2008), pp. 170—187.
133. Guyatt G., Walter S., Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis*, 40 (1987), pp. 171—178.
134. Leidy N.K., Revicki D.A., Geneste B. Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value Health*, 2 (1999), pp. 113—127.
135. Revicki D., Hays R.D., Cella D., et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*, 61 (2008), pp. 102—109.
136. Griebisch I., Coast J., Brown J. Quality-adjusted life-years lack quality in pediatric care: a critical review of published cost-utility studies in child health. *Pediatrics*, 115 (2005), pp. e600—e614.
137. Prosser L.A., Hammit J.K., Keren R. Measuring health preferences for use in cost-utility and cost-benefit analyses of interventions in children: theoretical and methodological considerations. *Pharmacoeconomics*, 25 (2007), pp. 713—726.
138. Ungar W.J. Challenges in health state valuation in paediatric economic evaluation: are QALYs contraindicated?. *Pharmacoeconomics*, 29 (2011), pp. 641—652.
139. Wright D.R., Wittenberg E., Swan J.S., et al. Methods for measuring temporary health states for cost-utility analyses. *Pharmacoeconomics*, 27 (2009), pp. 713—723.
140. Kromm S.K., Bethell J., Kraglund E., et al. Characteristics and quality of pediatric cost-utility analyses. *Qual Life Res*, 21 (2012), pp. 1315—1325.
141. Ratcliffe J., Flynn T., Terlich F., et al. Developing adolescent-specific health state values for economic evaluation: an application of profile case best-worst scaling to the Child Health Utility 9D. *Pharmacoeconomics*, 30 (2012), pp. 713—727.
142. Ravens-Sieberer U., Wille N., Badia X., et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Qual Life Res*, 19 (2010), pp. 887—897.
143. Stevens K. Valuation of the Child Health Utility 9D Index. *Pharmacoeconomics*, 30 (2012), pp. 729—747.
144. Wille N., Badia X., Bonsel G., et al. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. *Qual Life Res*, 19 (2010), pp. 875—886.
145. Prosser L.A. Comparative effectiveness and child health. *Pharmacoeconomics*, 30 (2012), pp. 637—645.
146. Shaddy R.E., Denne S.C. The Committee on Drugs and Committee on Pediatric Research. Clinical report-guidelines for the ethical conduct of studies to evaluate drugs in pediatric populations. *Pediatrics*, 125 (2010), pp. 850—860.
147. Shilling V., Williamson P.R., Hickey H., et al. Processes in recruitment to randomised controlled trials of medicines for children (RECRUIT): a qualitative study. *Health Technol Assess*, 15 (2011), pp. 1—116.

QALY: ОСНОВЫ

Milton C. Weinstein, PhD,¹ George Torrance, PhD,² Alistair McGuire, PhD³

¹ — *Harvard School of Public Health, Boston, MA, USA*

² — *McMaster University, Hamilton, ON, Canada*

³ — *London School of Economics, London, UK*

Переводчик:

Виллом Ирина Александровна — ассистент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И. П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Плавинский Святослав Леонидович — д.м.н., заведующий кафедрой педагогики, философии и права Северо-Западного Государственного Медицинского университета имени И. И. Мечникова, Санкт-Петербург, Российская Федерация
Белюсов Дмитрий Юрьевич — генеральный директор ООО «Центр фармакоэкономических исследований», г. Москва

Ключевые слова: откорректированная по качеству продолжительность жизни, анализ стоимости-эффективности, предпочтения, полезность, метод стандартной азартной игры, метод временной эквивалентности, типология, теория

QALYs: The Basics

Keywords: quality-adjusted life-year, cost-effectiveness analysis, preference, utility, standard gamble, time trade-off, taxonomy, theory

Факты

Целью данной работы является представить обзор концепции лет жизни, откорректированных на качество (quality-adjusted life-year (QALY)), широко используемого показателя изменений в состоянии здоровья, который используется при принятии решений о распределении ресурсов в здравоохранении. Исходно QALY были разработаны для измерения эффективности вмешательств в рамках анализа стоимости-эффективности, метода, который должен был помочь лицам, принимающим решения в области распределения ограниченных ресурсов между конкурирующими программами в области здравоохранения [1—3]. Далее исходная концепция QALY, в той форме, в которой она была определена в ранней литературе, будет называться «традиционная» модель QALY, признавая, что были предложены и альтернативные концептуальные модели, включая (но не ограничиваясь) так называемым «QALY с учетом весовых коэффициентов равенства и справедливости». Американский комитет по стоимости-эффективности в медицине и здравоохранении (US Panel on Cost-Effectiveness in Health and Medicine) [4] и британский Национальный институт здоровья и клинического совершенства (National Institute of Health and Clinical Excellence (NICE)) признали традиционную модель QALY в качестве «базового метода», т.е. стандартизованного методологического подхода, исполь-

зующегося для обеспечения сравнимости результатов анализа стоимости-эффективности различных вмешательств в области здравоохранения.

При использовании QALY предполагается, что основной задачей лиц, принимающих решения, является достижение максимального количества здоровья или улучшения здоровья в условиях бюджетных ограничений. Использование QALY далее предполагает, что здоровье и улучшение здоровья могут быть измерены или оценены на основе времени, которое индивид проводит в том или ином состоянии здоровья. По этой причине традиционное QALY измеряет выгоды, связанные со здоровьем. Однако мы осознаем, что лицо, принимающее решения может иметь и другие задачи, такие как равенство, справедливость и политические цели, которые в настоящий момент должны анализироваться вне традиционных моделей стоимости-эффективности, базирующихся на QALY.

Следующая работа в этом Специальном выпуске [5] анализирует некоторые из этих вариантов традиционной модели QALY. Исходно модель QALY не была разработана для того, чтобы помогать в принятии решений относительно индивидуальных пациентов, однако иногда она применялась и в клиническом анализе принятия решений именно для таких целей.

Основные идеи модели традиционного QALY основываются на теории принятия решений и теории

ожидаемой полезности. Основной идеей является то, что индивиды со временем проходят через различные состояния здоровья и каждое состояние здоровья имеет некоторую ценность. Здоровье, то что мы пытаемся максимизировать, определяется как время, умноженное на весовой коэффициент ценности — года жизни, откорректированные на качество — которое аккумулируется на протяжении определенного, разумного промежутка времени, что и дает нам QALY. Состояния здоровья должны оцениваться таким образом, чтобы ценность быть мертвым равнялась бы 0, поскольку отсутствие жизни считается 0 QALY. По договоренности, верхняя граница шкалы, определяемая как полное здоровье, имеет значение 1. Для того, чтобы значения QALY можно было бы складывать, шкала ценности должна быть интервальной, такой, чтобы выигрыш от 0,2 до 0,4 был бы так же ценен, как и выигрыш от 0,6 до 0,8. Состояния хуже смерти могут существовать, и они будут иметь отрицательные значения, которые будут вычитаться из количества QALY. Эти состояния, вместе с допущением нейтральности к риску на протяжении анализируемых лет жизни достаточны для того, чтобы модель QALY была полезным отражением предпочтений в отношении состояний здоровья.

Что такое цена?

В традиционной модели QALY состояние здоровья, которое является более привлекательным, считается более ценным. Цена, таким образом, уравновешивается с предпочтениями или желательностью. Основным вопросом тогда становится, желательным для кого?

Одной возможностью определить желательность состояния здоровья является определить, насколько индивид оценил бы свое пребывание в этом состоянии. В этом случае измеряются индивидуальные предпочтения, что делается при помощи опросов по предпочтениям, в которых для оценки предпочтительности различных состояний здоровья используются такие методы, как стандартная азартная игра, метод временной эквивалентности или визуальные аналоговые шкалы. Ключевым вопросом, к которому мы еще вернемся, является то, являются ли для принятия решения о распределении ресурсов важными оценки ценности лицами, которые в настоящий момент находятся в изучаемом состоянии, или тех, по заданию которых принимается решение и которые могут быть, а могут и не быть в оцениваемом состоянии здоровья на момент оценки. Альтернативой для оценки предпочтений состояний здоровья напрямую является оценка предпочтений небольшого набора доменов здоровья, или атрибутов, а затем построение мультиатрибутного показателя полезности как суммарного значения, которое отражает предпочтения как внутри, так и между доменами здоровья.

Как отмечено выше, предпочтительность состояния здоровья может быть также определена с точки зрения предпочтительности состояния здоровья населения в целом (т.е., в основном других людей, хотя, возможно, и включая респондента), а не исключительно самого человека. Оценка состояния здоровья населения в целом, вместо суммирования индивидуальных предпочтений того или иного состояния здоровья, позволяет лицу, принимающему решение учитывать при анализе и другие задачи, кроме максимизации здоровья. Мы вернемся к этому вопросу позднее, и он также обсуждается в другой статье этого Специального выпуска [5]. Предпочтения относительно состояния здоровья населения обычно измеряются таким методом, как метод эквивалентности жизней (person trade-off), хотя также могут использоваться методы стандартной азартной игры (standard gamble) временной эквивалентности (time trade-off) и ранговые (рейтинговые) шкалы (rating scales), иными словами все те методы, которые используются и при оценке предпочтений состояний здоровья с точки зрения самого респондента.

Важным аспектом традиционной модели QALY — вне зависимости от используемого метода изучения предпочтений — является то, что измеряются предпочтения относительно самого состояния здоровья, а не переходов из одного состояния в другое. Данный Специальный выпуск [5] (статья *Nord и соавт.*) рассматривает альтернативы традиционной модели QALY, которые оценивают изменения здоровья, а не его абсолютные уровни, как показатель ценности.

Как QALY используются?

Кого считать лицами, принимающими решения и что им надо? Некоторые широкие категории подходов к применению QALY при принятии решений представлены в столбцах табл. 1, и мы их сейчас обсудим.

Затем мы обратимся к различным особенностям процедур, которые используются для определения QALY, представленным в строках табл. 1. Во-первых, ценность чего изучается? Во-вторых, кого необходимо спрашивать, и чьи предпочтения важны: лиц, находящихся в данном состоянии здоровья сейчас, или других лиц? В-третьих, что надо спрашивать — какая методика определения ценности должна использоваться и как? В-четвертых, как определяются те исходы для здоровья, которые являются объектом оценки предпочтений? Это состояния здоровья (обычный подход), варианты динамики изменения состояний здоровья во времени или изменения состояния здоровья, либо в определенный момент, либо во времени? Мы последовательно обсудим эти аспекты QALY. Последние две строки табл. 1, к которым мы будем возвращаться при обсуждении каждого аспекта QALY указывают на дополнительные

Матрица, характеризующая подходы к использованию и определению QALY

Вопрос:	Решение индивидуального клинического вопроса или страхового покрытия		Общественный аудит: оценка идущих программ и мероприятий		Распределение ресурсов в обществе: распределение приоритетов между предлагаемыми программами и изменениями в них	
	Концептуализация ценности: чьи исходы с точки зрения здоровья и чьи предпочтения	Здоровье индивида: предпочтительность с точки зрения индивида до наступления состояния		Здоровье индивида: что человек чувствует, затем суммирование		Здоровье индивида: предпочтительность с точки зрения индивида до наступления состояния, затем суммирование
Кого спрашивают:	Индивида, информацию которому дает пациент/ инвалид†		Лиц, на которых направлено мероприятие, т.е. пациенты/инвалиды, те у кого «предотвращено» заболевание и т.д.		Репрезентативную популяционную выборку	Репрезентативную популяционную выборку
Методы оценки:	SG, ‡TTO, § RS		SG, ‡TTO, § RS, или инструменты MAU		SG, ‡TTO, § RS, или инструменты MAU	PTO, или трансформированные значения MAU
Исходы:	Полные профили здоровья на протяжении времени	Состояния здоровья и их длительность	Полные профили здоровья на протяжении времени	Состояния здоровья и их длительность	Состояния здоровья и их продолжительность (обычное применение QALY)	Состояния здоровья и их продолжительность
Необходимые дополнительные допущения	Никаких, если используется SG	Риск-нейтральность в отношении продолжительности жизни, суммируемость оценок разных временных периодов	Суммируемость оценок отдельных участников	Риск-нейтральность в отношении продолжительности жизни, суммируемость оценок разных временных периодов и разных участников	Риск-нейтральность в отношении продолжительности жизни, суммируемость оценок разных временных периодов и разных участников	Риск-нейтральность в отношении продолжительности жизни, суммируемость оценок разных временных периодов и разных участников
Дополнительные вопросы:			Равенство реальных исходов	Равенство реальных исходов	Равенство предполагаемых исходов	Равенство/справедливость, но в какой степени?

SG, стандартная азартная игра; TTO, временная эквивалентность; RS, рейтинговые шкалы; MAU, многоатрибутная полезность; PTO, равенство жизней.

* Оценивая изменение здоровья других, респонденты могут учитывать вопросы равенства и справедливости.

† Пациенты/инвалиды могут предоставить ценную информацию о величине выигрыша в результате адаптации, если она происходит, так, чтобы уменьшить степень снижения полезности в сравнении с опросами тех, кто не испытывал на себе данное состояние. Это уменьшение снижения полезности включает также и переоценку потерянных возможностей.

‡ SG может являться предпочтительным инструментом, поскольку только этот метод продемонстрировал интервальность шкалы предпочтений, т.е. изменения от 0,2 до 0,4 одинаково предпочтительны в сравнении с изменениями от 0,6 до 0,8 [8].

§ Теоретически оценки TTO эквивалентны оценкам SG при выполнении условий, которые должны выполняться, когда QALY используется для измерения полезности (риск-нейтральность в отношении продолжительности жизни).

|| Обратите внимание, что этот метод быстро теряет свою значимость ввиду большого числа возможных профилей здоровья.

допущения, которые необходимо принять для того, чтобы считать QALY адекватным показателем для измерения здоровья и улучшения здоровья, а также ряд дополнительных вопросов, которые игнорируются в QALY.

Каков основной вопрос?

Мы искусственно разделили все возможные вопросы, для ответов на которые может использоваться QALY, в широком понимании этой модели, на три основные категории, представленные в столбцах табл. 1. Первая категория (правый столбец) включает в себя вопросы распределения общественных ресурсов — иными словами установление приоритетов среди различных программ и возможных изменений в программах. Как указывалось, выше, это вопрос, для ответа на который, в первую очередь, и была разработана традиционная модель QALY. Вторая категория (левый столбец) отражает индивидуальные решения, которые принимает индивид в отношении своего собственного здоровья, включая вопросы лечения или выбора страховой программы. Третьей категорией, в которой может использоваться QALY, является то, что мы назвали общественным аудитом или программным аудитом (средний столбец), процедура оценки действующих мероприятий или программ с точки зрения здоровья населения. Для этой задачи необходимо располагать описанием здоровья населения, либо в определенный момент времени, либо его изменения со временем.

Последние две строки таблицы 1, к которым мы будем возвращаться при оценке каждого аспекта модели QALY, суммируют некоторые дополнительные допущения, выполнение которых необходимо, чтобы эта модель давала адекватный показатель состояния здоровья или улучшения здоровья, а также ряд других факторов, которые проигнорированы в модели QALY.

Что подвергается оценке?

Если речь идет о личных клинических решениях или решениях в области страхования, лицом, принимающим решение, является индивид (или домохозяйство), кто и заинтересован в оценке желанности каждого из возможных состояний здоровья. Мы часто подходим к этим решениям аналитически, используя деревья решений или различные оптимизационные модели, в которых лицо, принимающее решения, пытается оптимизировать ожидаемую полезность (utility). Поскольку точка зрения, с которой рассматривается проблема — это индивид, принимающий решение, используемые полезности — это полезности с точки зрения этого индивида в момент принятия решения. Поэтому, если возможными результатами принятого решения будут состояния

здоровья, которые человек никогда не испытывал, соответствующие предпочтения будут те, что имеются до наступления исхода (*ex ante*) и до появления любого опыта данного состояния. Конечно, разумный человек перед принятием решения будет стараться получить как можно больше информации об этих возможных состояниях здоровья, включая, возможно, опрос тех, кто имеет опыт нахождения в данном состоянии и может описать свои чувства в этом состоянии.

Для целей общественного или программного аудита, мы обычно заинтересованы в оценке нынешнего состояния здоровья лиц, у которых встречается изучаемое состояние или они охвачены программой, причем с их точки зрения. Альтернативным подходом является привлечь во внимание не только желательность нынешнего состояния здоровья, но и желательность будущих возможных состояний здоровья этих лиц, включая продолжительность жизни и предполагаемый прогноз для состояния здоровья. Основанием этого последнего подхода к общественному аудиту может быть то, что, если в популяции имеется широко встречающийся фактор риска будущей заболеваемости и смертности, их здоровье должно расцениваться как менее желательное, чем в популяции с аналогичным текущим уровнем здоровья, но лучшим прогнозом.

В то время, как индивидуальные решения относительно клинической тактики или страхования находятся под воздействием предпочтений состояний здоровья до получения опыта пребывания в них (*ex ante*), а общественный или программный аудит осуществляется путём измерения индивидуальных предпочтений у лиц, имеющих опыт нахождения в соответствующем состоянии (*ex post*), общественные решения о распределении ресурсов могут опираться на измерение ценности либо с точки зрения индивида (*ex ante* или *ex post*) или с точки зрения общественности. Для случая использования точки зрения индивидуумов это означает измерение желательности отдельных состояний здоровья для конкретного человека — как в случае индивидуальных решений — а затем суммирования оценок различных лиц. Этот подход с точки зрения конкретного человека совместим с принципом суверенитета потребителя, ключевого положения экономики благосостояния (welfare economics), и именно этот подход чаще всего используется в традиционной модели QALY. Индивидуальные предпочтения измеряются такими методами, как стандартная азартная игра, метод временной эквивалентности и визуальной аналоговой шкалы. Предпочтения, на которые опирается традиционная модель QALY — обычно формулируется до наступления исхода у респондента (*ex ante*), однако может опираться и на оценки тех, кто уже находится в соответствующем состоянии [6].

Для решений о распределении ресурсов с точки зрения общества, ценности, ассоциированные с тем или иным исходом для здоровья, могут быть получены путем выяснения того, что респондент чувствует относительно здоровья других или населения в целом, а затем оценки отдельных респондентов суммированы. Желательность состояний здоровья или изменений состояния здоровья других может включать и оценку справедливости каждым индивидом. Например, респондент может оценивать выше изменения здоровья, если они анализируются для лиц, у которых исходно состояние здоровья было хуже, в сравнении с теми, кто исходно имеет более высокий уровень здоровья.

Кого мы спрашиваем?

С точки зрения общества, мы можем спрашивать о предпочтениях или ценности их собственного здоровья или о здоровье других. Для того, чтобы определить, как люди оценивают их собственные возможные состояния здоровья, необходимо опросить репрезентативную выборку пациентов с данным состоянием — эта выборка должна включать пациентов или инвалидов, которые встречаются в общей популяции — и получить у них информацию о предпочтениях определенных состояний. Лица, которые сами не находились в определенном состоянии в идеале должны получить информацию от пациентов и инвалидов, каково это, находиться в данном состоянии. Хотя можно провести различие между подходом *ex ante* и *ex post*, когда, в последнем случае, предпочтения получаются только от лиц, которые в настоящий момент находятся в изучаемом состоянии [6], это не вполне взаимоисключающие подходы. Респонденты могут получить информацию об адаптации к определенным состояниям здоровья, от тех, кто их испытывает, когда их просят оценить эти состояния здоровья.

Несколько иным способом ответа на вопрос, о том, кого спрашивать о предпочтениях, которые могли бы использоваться при принятии решений о распределении общественных ресурсов, является предложение оценить предпочтения, связанные со здоровьем населения в целом, а не индивидуальные предпочтения, связанные с собственным здоровьем. При оценке здоровья других лиц, мы все равно, скорее всего, сформируем выборку из общей популяции, которая будет включать пациентов и инвалидов, с той частотой, с которой они в этой популяции встречаются, и респонденты будут информированы о том, какие ценности пациенты или инвалиды приписывают своему здоровью.

Для общественного или программного аудита мы обратимся к тем, кто представляет интересующую нас популяцию, а для личных клинических решений мы обратимся к самому пациенту, как к лицу, при-

нимающему решение. В контексте индивидуальных решений, очевидно, что мы заинтересованы в собственных предпочтениях респондента, с обеспечением того, чтобы эти предпочтения базировались на знаниях людей, находящихся в анализируемых состояниях здоровья, таких, собственного опыта пребывания в которых у респондента может быть и нет, но может появиться в будущем.

Какой вопрос мы задаем?

Стандартная азартная игра является методом выбора для личных решений относительно клинической тактики или страхования, она базируется на принципах максимизации ожидаемой полезности из теории принятия решений. Тем не менее, методы временной эквивалентности и, возможно, ранговые шкалы, также используются для того, чтобы оценить состояния здоровья ввиду опасений, что стандартная азартная игра может оказаться под влиянием когнитивных ошибок при ее проведении [7].

Для индивидуальных подходов при анализе общественного мнения, предпочтения для QALY оцениваются с помощью стандартной азартной игры, метода временной эквивалентности, ранговые шкалы или мультиатрибутивные инструменты, которые используют те же методы для получения предпочтений внутри и между определенных доменов здоровья. Для оценки здоровья других или населения в целом, дополнительной возможностью является метод эквивалентности жизней. Метод эквивалентности жизней может использоваться и на индивидуальном уровне, однако он более естественен при оценке здоровья населения.

Как мы задаем эти вопросы о ценности?

Метод стандартной азартной игры базируется на широко принятых аксиомах постоянства предпочтений в условиях неопределенности, таких как транзитивность, независимость и непрерывность. Стандартная азартная игра является единственным из перечисленных методов, имеющим свойства измерять предпочтения на интервальной шкале, иными словами изменения от 0,2 до 0,4 равны изменениям от 0,6 до 0,8 [8]. Метод временной эквивалентности дает результаты, примерно равные таковым при использовании стандартной азартной игры, как было продемонстрировано в нескольких эмпирических исследованиях. При этом данным методом имеет уникальные концептуальные взаимоотношения с QALY поскольку этот метод открыто приравнивает время, проведенное в состоянии сниженного здоровья и время (меньшее), проводимое в состоянии полного здоровья (время, откорректированное на качество). Результаты, получаемые при использовании метода временной эквивалентности теоретически эквива-

лентны результатам, получаемым при использовании стандартной азартной игры в условиях, когда можно пользоваться моделью QALY, допущения, которые включают риск-нейтральность в отношении продолжительности жизни. Ранговые шкалы (включая визуальные аналоговые шкалы) рассматриваются обычно как теоретически менее обоснованным методов в сравнении со стандартной азартной игрой или методом временной эквивалентности ввиду возможных смещений, связанных с определением их шкалы и фактом, что они требуют оценить ситуацию, а не выбор ситуации. Тем не менее, при их использовании не возникает когнитивных ошибок, как это может произойти при использовании вероятностей в стандартной азартной игре [7], а также они не подвержены временному дисконтированию как в случае метода временной эквивалентности [9].

Примерами мультатрибутных инструментов оценки полезности являются шкала EuroQOL 5-утверждений (EQoL-5D), шкала 7- утверждений Health Utilities Index 2, шкала 8- утверждений Health Utilities Index 3, шкала 6- утверждений SF-6D, базирующаяся на SF-36, шкала 4- утверждений Quality of Well-Being, шкала 15-утверждений 15D scale, и шкала 5-утверждений Assessment of Quality of Life. При использовании этого подхода, оцениваемые состояния здоровья представлены в виде матрицы, представляющей собой комбинацию доменов здоровья (или атрибутов), которые определяются тем или иным инструментом. Например, 243 состояния здоровья в EQ-5D определяются путем выбора одного из трех уровней здоровья в каждом из пяти доменов. Пациенты относят себя к одной из ячеек этой матрицы. Каждой ячейке соответствует показатель, который ранее был получен путем опроса популяционной выборки. Это метод выбора для большинства современных исследований стоимости-эффективности. Этот подход рекомендован Комиссией по стоимости-эффективности в медицине и здравоохранении [4] и его предпочитает NICE. Проблемой является то, что разные инструменты дают разные результаты, частично вследствие выбора различных атрибутов, частично благодаря тому, как моделируются взаимоотношения между отдельными атрибутами. Более того, оценки могут быть специфичными для определенной популяции (в зависимости от страны или социодемографических характеристик респондентов) однако эмпирически такая специфичность выявляется реже, чем можно было бы ожидать

Метод эквивалентности жизней является примером другого подхода, который мы могли бы назвать методами популяционной эквивалентности. В этом случае респонденты выступают в роли заместителей лиц, принимающих решения и принимают гипотетические решения в отношении конкурирующих вмешательств, как будто бы они были диктаторами, пытающимися творить добро. Например, в рамках

метода временной эквивалентности респондента могут попросить сделать выбор между помощью X пациентам с состоянием A и улучшением до уровня A' и помощью Y пациентам в состоянии B и улучшением до уровня B'. Парадигма метода эквивалентности жизней предполагает опрос репрезентативной выборки членов популяции на предмет сравнения желательности представления определенному числу пациентов различные вмешательства. В одном из вариантов этого подхода, X или Y фиксируется и второе число корректируется до тех пор, пока респонденты не оказываются в замешательстве, важность улучшения от B к B' в сравнении с улучшением от A к A'. Обратите внимание, что метод эквивалентности жизней базируется на основе решения о ценности состояний здоровья других. Другим важным фактором является то, что метод эквивалентности жизней позволяет респондентам учитывать некоторые аспекты равенства или справедливости, в частности в том, что ценность улучшения здоровья может зависеть от исходного уровня здоровья (A или B в нашем примере) и предоставить большую ценность улучшению здоровья тех, кто исходно был в более плохом состоянии.

Как дается определение исхода?

В традиционной модели QALY исход определяется как определенное состояние здоровья, которое фиксируется на данный момент времени. Затем полезности каждого состояния здоровья, умноженные на продолжительность пребывания в этом состоянии, суммируются и дают количество QALY, причем для анализа стоимости-эффективности к QALY применяется дисконтирование. За счет предположения о том, что полезность (utility) пребывания в некоем состоянии здоровья не зависит ни от времени, проведенном в этом состоянии, ни от последовательности состояний здоровья, предшествующих и следующих за анализируемым, время устраняется из процесса оценки полезностей. Критическим допущением тогда становится, что после получения величины QALY могут суммироваться во времени, возможно, с использованием весовых коэффициентов временных предпочтений, если применяется дисконтирование.

Более общим подходом, который расслабляет допущение о том, что значение состояния здоровья не зависит от состояний, предшествующих или следующих за изучаемым, ни от времени проведенном в этом состоянии (т.е. допущение межвременной независимости полезностей), является присвоение значение полезности последовательности состояний здоровья в зависимости от времени, которые называются профилями здоровья [10]. Оценка профилей здоровья является более общим подходом, чем оценка состояний здоровья с последующим суммированием оценок, и поэтому теоретически лучше,

поскольку не полагается на допущение возможности суммирования оценок на протяжении изучаемого промежутка времени. Однако практическая проблема с оценкой профилей здоровья заключается в том, что, поскольку исходы определяются как последовательность состояний здоровья на протяжении всей жизни, их число потенциально крайне велико. В контексте Марковской модели, симуляции на уровне пациента или даже клинического испытания с несколькими периодами наблюдения, задача оценки всех возможных профилей становится практически не решаемой [9]. Если имеется N возможных состояний здоровья и T временных периода, подход с использованием профилей здоровья требует оценки $N \cdot T$ вариантов.

Третий подход, описанный более детально в статье *Nord и соавт.* В этом специальном выпуске [5], не оценивает ни абсолютные состояния здоровья, ни профили, а акцентируется на изменениях в состоянии здоровья. Вместо того, чтобы оценивать состояние здоровья X и предпочтительное состояние здоровья Y отдельно, напрямую определяется ценность перехода из X в Y . Метод эквивалентности жизней в особенности фокусируется на исходном и достигнутом уровне здоровья и, если необходимо, может присвоить большую ценность улучшению здоровья с менее желательного исходного уровня. Одним из критических отличий этого подхода и традиционной модели QALY является то, что в сочетании с методом эквивалентности жизней, этот метод изучения предпочтений изменения здоровья акцентируется на здоровье популяционной группы, к которой может принадлежать (а может и не принадлежать) респондент.

При использовании этого подхода возникает то же ограничение, что и при использовании профилей здоровья: количество заданий по получению

предпочтений может быть очень большим. Если имеется N возможных состояний здоровья, данный подход требует $N \cdot (N-1)$ оценок — это число меньше, чем при использовании профилей здоровья, но все равно больше числа возможных состояний здоровья. Например, если желательно использовать такое количество состояний здоровья, как в EQ-5D (243 состояния здоровья), то необходимо будет оценить $243 \cdot 242 = 58806$ возможных изменений состояний здоровья. Еще одной проблемой данного подхода является то, что характер перехода из одного состояния здоровья в другое может изменить оценку такого перехода: ценность улучшения от X до Y , а затем от Y до Z может отличаться от ценности прямого улучшения от X до Z . На самом деле, если оценка изменений не зависит от того, как они происходят, метод дает такие же результаты, как и при оценке только самих состояний здоровья.

Основы концепции QALY

В заключение к настоящему очерку и в качестве введения для последующих статей определим 9 основных допущений, относящихся к традиционной модели QALY, как она используется при принятии решений о распределении общественных ресурсов (табл. 2).

Во-вторых, последствия для здоровья от принятия альтернативных решений могут быть описаны с точки зрения состояний здоровья, их изменений и продолжительности пребывания людей в каждом из состояний здоровья. Все последствия, не связанные со здоровьем, измеряются либо как экономические расходы и включаются в числитель отношения стоимости-эффективности, либо игнорируются в количественном анализе в рамках модели стоимости-эффективности.

Таблица 2

Основные принципы традиционной модели QALY

1.	Необходимо принять решение о распределении ресурсов.
2.	Результат принятия альтернативных решений может быть описан с точки зрения состояний здоровья, их изменений и продолжительности пребывания в них пациента.
3.	Ресурсы ограничены, и каждое альтернативное решение потребляет ресурсы (есть расходы).
4.	Основной задачей лица, принимающего решение, является максимизация здоровья населения в условиях ограниченности ресурсов.
5.	Здоровье определяется как время, откорректированное на качество (ценность), проанализированное за разумный временной промежуток.
6.	Качество (ценность) измеряется с точки зрения предпочтения (желательности).
7.	Каждый индивид нейтрален к риску с точки зрения продолжительности жизни и полезности могут суммироваться во времени.
8.	Оценки ценностей (предпочтений) измеренные у разных респондентов могут суммироваться и использоваться для группы в целом.
9.	QALY могут быть суммированы для разных пациентов, иными словами один QALY — это один QALY, вне зависимости от того, кто его приобретет или потеряет.

QALY, качество жизни с поправкой на год.

В-третьих, ресурсы ограничены, поэтому каждое альтернативное решение оказывает воздействие на доступность ресурсов, иными словами имеются альтернативные издержки.

В-четвертых, основной задачей лица, принимающего решения является максимизация здоровья населения в условиях ограниченного количества ресурсов.

В-пятых, здоровье определяется как время, откорректированное на качество жизни (его ценность), на протяжении разумного временного периода.

В-шестых, качество жизни (его ценность) измеряется с точки зрения его желательности или предпочтительности.

Эти шесть требований не ограничивают аналитика с точки зрения метода, используемого для спецификации или описания исходов для здоровья, которые затем применяются для расчета QALY. Далее мы обратимся к более жестким (и дискутабельным) допущениям, которые относятся именно к модели традиционного QALY.

В-седьмых, каждый индивид является нейтральным к риску с точки зрения продолжительности жизни и полезности могут суммироваться во времени. Нейтральность риска необходима для того, чтобы обосновать возможность расчета откорректированной на качество продолжительности жизни, иными словами среднюю ценность возможного числа QALY, причем каждое откорректировано на вероятность появления. Возможность суммирования во времени, обсуждавшаяся ранее, это допущение, позволяющее нам предпринимать оценку только состояний здоровья в определенный промежуток времени, игнорируя их длительность и последовательность. Это очень сильные допущения по поводу предпочтений, которые, без сомнения, упрощают реальность, но они необходимы для того, чтобы QALY описали функцию ценности здоровья для индивида на протяжении времени. Сказать, что эмпирические данные противоречивы в отношении того, насколько данные допущения приводят к разумной аппроксимации реальности, наверное, будет не сказать ничего. В большинстве случаев имеются доказательства того, что большинство людей не принимают риск, когда речь заходит о продолжительности жизни (risk-averse), хотя, когда они действуют как представители общества, это немного не так. В то же время значительная часть литературы говорит, что допущение возможности суммирования во времени может быть истинным, а может и нет. Например, есть данные, что люди могут существовать с проблемой со здоровьем на протяжении короткого периода времени, а ощущаемое влияние на здоровье тем сильнее, чем дольше они его имеют. Этот феномен называется «максимальное время выдержки» [11]. С другой стороны, имеются доказательства тому, что люди могут адаптироваться к негативным состояниям здоровья.

Любая из этих характеристик нарушает допущение возможности суммирования полезностей во времени, хотя и разнонаправленно.

В-восьмых, ценности или предпочтения, измененные у разных людей, могут суммироваться и использоваться для группы в целом.

Наконец, QALY, рассчитанные с использованием усредненных весовых коэффициентов предпочтений могут суммироваться для разных пациентов.

Другие вопросы, связанные с QALY: дисконтирование и весовые коэффициенты равенства

Суммарная оценка QALY, как признанный результат вмешательства при принятии решения о распределении общественных ресурсов в здравоохранении, как и в других областях, должна отражать положительное предпочтение ближайшего времени или дисконтирование. Дисконтирование как расходов, так и QALY, причем по одной и той же ставке, в настоящий момент является стандартным допущением и рекомендуется как американским Советом, так и британским NICE. Несмотря на сближение точек зрения в этих двух странах, имеются спорные взгляды относительно дисконтирования. В некоторых странах, например, Голландии, существует точка зрения, что расходы должны дисконтироваться по более высокой ставке, чем QALY. Кроме того, есть точка зрения, что QALY вообще не должны дисконтироваться, если выполняется допущение нейтральности к риску. Аргумент за дисконтирование QALY базируется, в основном, на концепции альтернативных издержек, как описано, например, американским Советом [4], но при этом сохраняется ключевое допущение о том, что люди, как индивиды, нейтральны к риску QALY во времени оцениваются одинаково. Подобная проблема продолжает существовать и, пока что, полностью не решена.

Вопросы равенства и справедливости не включены количественно в модель традиционного QALY, за исключением основного допущения, что каждый QALY для каждого пациента имеет одинаковый весовой коэффициент. Это не означает, что данные вопросы не важны, но должны учитываться лицом, принимающим решения, вместе с суммарными значениями выигрыша QALY. Суммарные выигрыши в здоровье, измеренные при помощи QALY — один из многочисленных факторов, которые используются в процессе принятия решений в клинике, общественном и программном аудите или при принятии решений о распределении ресурсов. Другие факторы, включая равенство и справедливость, должны рассматриваться, в традиционной модели QALY, отдельно. Примерами некоторых дополнительных аспектов является возраст целевой популяции, базовое состояние здоровья, и, возможно, принцип,

согласно которому более ценным является повысить минимальный уровень здоровья населения, нежели повысить средний уровень за счет улучшения состояния здоровья и так достаточно здоровых лиц. Политические условия также могут оказывать воздействие на приоритеты лица, принимающего решения, в отношении распределения ресурсов. Может ощущаться потребность приоритезировать помощь пациентам с редкими заболеваниями, лечить конкретного пациента в противоположность профилактике статистического заболевания, или в обеспечении ухода самым маленьким и пожилым, поскольку они рассматриваются как самые уязвимые.

QALY принимают важное участие в принятии решений в области здравоохранения. С оговорками, отмеченными выше, традиционная модель QALY остается серьезным концептуальным инструментом, который, как мы убеждены, может привести к

улучшению принятия решений. QALY помогают делать выбор, но имеются многочисленные иные условия при принятии решений в здравоохранении, которые не учитываются в традиционной модели QALY. Тем не менее, традиционная модель QALY и не разрабатывалась, как включающий все проблемы инструмент для лиц, принимающих решения [4]. Учитывая все допущения и сложности, связанные с измерением, важно продолжать использовать QALY с осторожностью. Однако мы придерживаемся точки зрения, что традиционная модель QALY продолжает играть важную роль в принятии решений в здравоохранении.

Благодарности

Авторы искренне благодарны следующим членам ISPOR: *Norman Daniels*, *Mark Kamlet*, и *Erik Nord*.

Литература

1. *Fanshel S., Bush J.W.* A health status index and its application to health services outcomes. *Oper Res* 1970;18:1021—66.
2. *Torrance G.W., Thomas W.H., Sackett D.L.* A utility maximization model for evaluation of health care programs. *Health Serv Res* 1972;7:118—33.
3. *Weinstein M.C., Stason W.B.* Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977;296:716—21.
4. *Gold M.R., Siegel J.E., Russell L.B., Weinstein M.C., eds.* Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
5. *Nord E., Daniels N., Kamlet M.* QALYs: some challenges. *Value Health* 2009;12(Suppl. 1):S10—5.
6. *Dolan P., Kahneman D.* Interpretations of utility and their implications for the valuation of health. *The Economic Journal* 2008;118:215—34.
7. *Bleichrodt H., Pinto J.L., Wakker P.P.* Making descriptive use of prospect theory to improve the prescriptive use of expected utility. *Manag Sci* 2001;47:1498—514.
8. *G.W.* Utility measurement in health care: the things I never got to. *Pharmacoeconomics* 2006;24:1069—78.
9. *Johannesson M., Pliskin J.S., Weinstein M.C.* Are healthy-years equivalents an improvement over quality-adjusted life years? *Med Decis Making* 1993;13:281—6.
10. *Mehrez A., Gafni A.* Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med Decis Making* 1989;9:142—9.
11. *Sutherland H.J., Llewellyn-Thomas H., Boyd N., Till J.E.* Attitudes toward quality of survival—the concept of “maximal endurable time.” *Med Decis Making* 1982;2:299—309.

Сохранить и улучшить QALY

Joseph Lipscomb, PhD¹, Michael Drummond, PhD², Dennis Fryback, PhD³, Marthe Gold, MD, MPH⁴, Dennis Revicki, PhD⁵

¹ — Department of Health Policy and Management, Rollins School of Public Health, Emory University, Atlanta, GA, USA

² — University of York, York, UK

³ — University of Wisconsin, Madison, WI, USA

⁴ — City University of New York Medical School, New York City, NY, USA

⁵ — United BioSource Corp, Bethesda, MD, USA

Переводчик:

Виллом Ирина Александровна — ассистент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И. П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Плавинский Святослав Леонидович — д.м.н., заведующий кафедрой педагогики, философии и права Северо-Западного Государственного Медицинского университета имени И.И. Мечникова, Санкт-Петербург, Российская Федерация

Белоусов Дмитрий Юрьевич — генеральный директор ООО «Центр фармакоэкономических исследований», г. Москва

Ключевые слова: годы жизни, откорректированные на качество, связанное со здоровьем качество жизни, анализ цена-польза, оценка исходов (в здравоохранении), статус здоровья

Retaining, and Enhancing, the QALY

Keywords: quality-adjusted life-years, health-related quality of life, cost-utility analysis, outcome assessment (health care), health status

Введение и обзор

Среди средств экономического и политического анализа, требующего интегральной оценки исходов, с учетом как количества лет жизни, так и ее качества, широкое применение получил показатель QALY (годы жизни, откорректированные на качество) [1–3].

Американская комиссия по экономической эффективности в медицине и здравоохранении (US Panel on Cost-Effectiveness in Health and Medicine) рекомендовала QALY для использования в «базовых» ситуациях [2]. Британский Национальный Институт Здравоохранения и Клинического Совершенствования (National Institute for Health and Clinical Excellence, NICE) регулярно использует QALY при оценке медицинских технологий [3]. Агентства по надзору и закупкам в сфере здравоохранения ряда других стран Европы, в Канаде и Австралии также признают QALY пригодным для анализа экономической эффективности применения тех или иных продуктов или мероприятий [4]. Энтузиазм, с которым академические аналитики и исследователи-эконо-

мисты используют QALY, становится очевидным после беглого обзора главных журналов по политике в сфере здравоохранения и программы дискуссий любой из последних ежегодных международных конференций ISPOR [5].

Прежде чем начать разговор, необходимо ввести понятие традиционного QALY:

$$QALY_{conv} = \sum_{t=1...T} \sum_{s=1...S} p_{st} V(H_{st})(1+r)^{t-1}$$

где p_{st} — вероятность, с которой пациент будет находиться в состоянии H_{st} в момент времени t ; $V(H_{st})$ — показатель ценности (или предпочтения), которая ассоциирована с пребыванием лица в состоянии H_s в момент времени t ; $(1+r)^{t-1}$ поправочный коэффициент (коэффициент дисконтирования), предназначенный для перевода полученной $V(H_{st})$ к текущим ценам, где r — коэффициент дисконтирования, отражающий временные предпочтения по исходам; S — число дискретных состояний здоровья, в которых может находиться лицо; T — временной горизонт, который важен для принятия решения; и наличие нижнего индекса у QALY в форме $QALY_{conv}$

следует трактовать, как то, что это (некий вариант) «традиционного» показателя QALY. Обратите внимание, что, когда QALY используется для принятия решения с точки зрения группы лиц (чаще всего при анализе с точки зрения общества), $V(\cdot)$, как правило, предназначена для отражения среднего арифметического предпочтений всех членов группы.

Как будет показано далее, большинство споров о QALY и большая часть возможностей по укреплению данного интегрального параметра для оценки исходов, лежит в области определения и классификации состояний здоровья (H_{st}), их оценки (V) и адекватности $QALY_{conv}$ в качестве руководства для разработки решений, поскольку приведенная выше формула не учитывает распределительные, этические и прочие важные аспекты. Разумеется, первые трудности встречаются уже на этапе сбора и анализа клинико-эпидемиологических данных для получения статистически значимых результатов оценки вероятностей нахождения в разных состояниях здоровья (p_{st}), однако в центре внимания настоящей статьи находятся определения этих состояний, их оценка и вопрос распределения ресурсов (так называемый «Q»-компонент QALY).

Показатель QALY, в соответствии с данным выше определением, представляет собой один из двух основных компонент анализа стоимости-полезности программ здравоохранения, который направлен на отбор вмешательств, применение которых приведет к дополнительному приросту QALY с наименьшими материальными затратами (другими словами, наибольший дополнительный прирост единиц QALY на каждый затраченный доллар) [1]. Более того, варианты компонента H_{st} традиционного QALY используются изолированно для отслеживания состояния здоровья населения на национальном [6, 7] или даже на локальном уровне [8].

Проблема №1: Сохраняющиеся различия во взглядах научной общественности на ряд ключевых технических и методологических аспектов расчета QALY

Вопросы, которые необходимо учитывать при использовании показателя QALY включают: 1) отбор специфических доменов или граней здоровья, поскольку в целом здоровье многогранно; 2) отбор специфических вопросов, которые позволяют составить представление о всём домене; 3) психометрические подходы для выявления индивидуальных предпочтений в отношении определенных состояний здоровья; 4) статистическое моделирование для получения взвешенных показателей ценностей, используемых при расчете QALY. В приведенном выше уравнении первым двум пунктам соответствует показатель H_{st} , а 3 и 4 связаны с $V(\cdot)$ и тем, как они со-

вместно позволяют отразить совокупность показателей здоровья в шкале QALY.

Различия в трактовке перечисленных вопросов на практике могут быть прослежены на примере тех систем измерения качества жизни, связанного со здоровьем (Health-Related Quality of Life, HRQOL), которые наиболее часто используются для получения QALY. К этим системам относятся — Индекс полезности здоровья Mark2 (Health Utilities Index, HUI2) и Mark 3 (HUI 3) [9], EuroQoLEQ-5D [10], индекс благополучия (QWB) [11], SF-6D [12], а также индекс ограничения здоровья и трудоспособности (HALex) [13]. Как показал Fryback и соавт. [7], данные шесть индексов отражают «сходные, но не идентичные тенденции» в HRQoL для старших возрастных групп населения США. Следовательно, будучи использованы одновременно для оценки экономической эффективности того или иного вмешательства, эти инструменты могут дать различные результаты по шкале QALY и, таким образом, сформировать различные представления о соотношении цена/польза данного вмешательства. На практике оказалось проще рекомендовать использовать QALY в качестве полезного общего подхода для оценки суммарного эффекта на здоровье, чем указать, какой именно подход должен использоваться в качестве стандартного.

Проблема №2: Стандартный QALY незаметно принимает допущения в отношении одних важных вопросов и просто игнорирует другие

Критику с этой стороны формально можно разделить на три общих категории, которые, в основном, относятся к тому, как определяется ценность тех или состояний здоровья. Во-первых, даже сторонники стандартного QALY признают, что в расчетах применяются упрощающие допущения, которые, как минимум, требуют тщательного анализа [14]. Во-вторых, утверждается, что в концепции QALY игнорируется ряд аспектов социальной справедливости и равенства, которые необходимо принимать во внимание в разработке решений для общества (при анализе цена/полезность) [15, 16]. В-третьих, в основе обычной модели расчетов QALY и её вариантов лежит концепция предпочтений, хотя иногда имеются различия во мнениях, чьи предпочтения необходимо учитывать и как. Однако, *Dolan u Kahneman* [17] и *Hausman* [18] возражают (по различным причинам) против самой идеи, что оценка состояний здоровья должна базироваться на индивидуальных предпочтениях. Данный круг вопросов будет кратко рассмотрен ниже.

Проблема №3: Концептуальные и методологические проблемы могут препятствовать внедрению методологии QALY

(и анализа стоимости-эффективности в целом) в ту область принятия решения, для которой они и разрабатывались

QALY играл и продолжает играть существенную роль в вопросах регуляции и принятия решений о закупках во множестве юрисдикций за пределами США. Более того, Служба Управления и Бюджета США сегодня требует от федеральных агентств дополнять анализ экономической выгоды анализом экономической эффективности с использованием обобщенного варианта QALY (продолжительность жизни, скорректированная по качеству) для принятия «экономически значимых решений в сфере охраны здоровья и безопасности» [19]. Однако в целом, анализ экономической эффективности, в том числе с помощью QALY, не имеет широкого распространения в практике разработки решений в США (см. *Neumann* [20]). Особо следует отметить, что Управление по лекарствам и пищевым продуктам США (FDA) не приветствовало использование основанных на предпочтениях методов оценки важных для пациентов исходов в контексте регистрации лекарственных препаратов [21]. Американские Центры обслуживания по программам Medicare и Medicaid до сих пор формально не используют в своей практике анализ стоимости-эффективности для принятия решений по страховому покрытию и возмещению [22]. Хотя анализ экономической эффективности с помощью QALY может служить тонким механизмом политического регулирования [20], пока остаётся неясным, какая доля противников использования QALY имеет претензии к его методологии, а какая — к историческим и политическим сложностям, связанным с открытым рacionamento в здравоохранении.

Выявив глубокие противоречия, которые таит в себе модель QALY, можно, казалось бы, заключить, что пора постепенно отвыкать или даже отказываться от данной интегральной системы в пользу альтернативных методов оценки значения для здоровья различных вмешательств. Мы считаем подобную позицию неблагоприятной в связи с доводами, которые приведем ниже. Это, в свою очередь, не означает «слепого» принятия традиционного QALY в качестве «золотого стандарта». Скорее, продолжающиеся исследования и эксперименты с QALY для совершенствования самой модели и оценки исходов в целом должны базироваться на четком понимании сильных и слабых сторон существующих подходов. Наиболее ярким и заметным среди них и является традиционный QALY.

Модель традиционного QALY оказалась эффективным средством количественной совместной оценки заболеваемости и смертности на протяжении времени, как с индивидуальных, так и с общественных позиций. За последние три десятилетия большие ин-

вестиции государственных и частных ресурсов были направлены на развитие, испытание и оценку базирующихся на предпочтениях показателей HRQoL, в том числе шести перечисленных выше. В комплексной оценке результатов применения EuroQol EQ-5D, QWB и HUI2/3 в онкологии, *Feeny* [23] пришёл к выводу, что все три системы подтвердили свою надежность, интерпретируемость, воспроизводимость и целесообразность. В Соединенных Штатах, финансируемый из федерального бюджета реестр опубликованных исследований анализа экономической эффективности продолжает отслеживать прогресс в том, насколько исследования придерживаются рекомендаций Комиссии США по эффективности затрат в здравоохранении и медицине [24]. Комитет Института медицины, который рекомендует Службе Управления и Бюджета США, использовать определенные показатели улучшения здоровья при принятии регуляторных решений рекомендовал то, что мы называем традиционным QALY как «наилучшую в настоящий момент меру стандартизированной оценки лет жизни, скорректированных на состояние здоровья, в связи с его широкой распространённостью, гибкостью и относительная простотой использования» [20].

Показатели HRQoL, которые предполагают применение модифицированных моделей QALY, теперь являются частью различных национальных баз данных в Канаде [6], США [7], и нескольких других странах [25]. Более того, появляется всё больше свидетельств того, что общественность в целом все чаще склонна учитывать эффективность затрат при принятии решений в здравоохранении. Недавнее исследование, проведенное *Bryan и соавт.* [26] демонстрирует, что в выборке лиц, принимающих решения, которая включала государственных и частных страховщиков, представителей медицинских организаций, основанных страховыми компаниями и работодателей в Калифорнии, распространено глубокое понимание необходимости определения приоритетов для эффективных решений. Участники фокус-группы выразили своё беспокойство по поводу возможного смещения в результатах спонсируемых фармацевтической индустрией исследований, а также возможностью судебных исков против организаций, которые первыми начинают внедрять анализ экономической эффективности в процесс принятия решений. Тем не менее, 90% участников дискуссии считали, что анализ стоимости-эффективности и QALYs важны для информированного принятия решений, более того, программа Medicare должна взять на себя инициативу в продвижении этой методологии.

Анализ сотен опубликованных применений традиционной модели QALY, дает значительную информацию о сильных и слабых сторонах в методологии данного основанного на предпочтениях оценочного

подхода. Отказаться от модели QALY в настоящий момент означает прервать связь с сотнями опубликованными исследованиями и многочисленными идущими научными работами — и, тем самым, сделать сложным, если не вообще невозможным, оценить насколько альтернативные показатели здоровья «улучшают» традиционную модель QALY. Аналогичным образом будет сильно скомпрометирована способность выполнять исторические сравнения экономических оценок или оценок состояния здоровья населения. Более продуктивным подходом явилось бы начало исследовательской программы, которая бы взяла традиционную модель QALY в качестве стартовой точки и затем попыталась бы справиться с проблемами, описанными выше

В последующих разделах мы будем опираться на данное обсуждение, и идентифицируем ряд тем, которые заслуживают пристального внимания. По отношению к самой модели QALY имеется три основных вопроса: определение и описание состояний здоровья, оценка этих состояний и подходы к дополнению QALY дополнительной информацией о социальном равенстве и вопросах перераспределения ресурсов. Без сомнения, имеются серьезные альтернативы для модели QALY, которые используются для оценки исходов в области здоровья, но они имеют свои методологические и практические проблемы. Поэтому мы приходим к выводу, что имеется потребность в дополнительных исследованиях в области «сравнительной эффективности», под чем мы понимаем исследования, которые критически сравнивают альтернативные подходы к измерению и оцениванию важных для здоровья исходов в области принятия решений в здравоохранении.

Определение и описание состояний здоровья

По мере своего возникновения и изменения во времени, основные системы классификации состояний здоровья, описанные ранее, сыграли полезную роль в экономической оценке вмешательств в области здравоохранения и в оценке здоровья населения.

Эти измерительные системы являлись «мотором», который лежал в основе многих (хотя и не всех) основанных на предпочтениях показателях, использовавшихся в исследованиях стоимость-полезность, опубликованных в последние годы. Тот факт, что эти системы, использующиеся для расчета традиционного QALY были применены в огромном числе экономических оценок для широкого спектра заболеваний и вмешательств, сам по себе не означает, что улучшать их не следует. Действительно, как отмечают *Brauer и соавт.* [24], большинство опубликованных в период с 1976 по 2001 года исследований, которые полагались на весовые коэффициенты полезностей не использовали общую систему измерения состояний здоровья, а применяли весовые коэффициенты

предпочтений, полученные специально для данного исследования. Вместе с тем, если посмотреть с другой стороны, эти системы измерения состояний здоровья прошли «тест рынком» в том, что многие аналитики (и лица, принимающие решения) согласны положиться на них при проведении экономической оценки в частном и государственном секторах.

Кроме того, вопросы из нескольких из этих измерительных систем были использованы в крупных скринингах состояний здоровья населения, давая возможность получить показатели популяционного здоровья, основанные на предпочтениях, и их изменение с течением времени. Самыми известными примерами являются:

1. Опрос по медицинским расходам США (US Medical Expenditure Panel Survey [27]), ранее включал как EuroQol EQ-5D, так и SF-12. В настоящий момент задаются вопросы только из SF-12, но недавно проведенный анализ показывает, как можно использовать ответы на вопросы SF-12 для оценки шкалы EQ-5D [28—30].
2. Как совместный опрос по здоровью Канады-США (Joint Canada—US Survey of Health [31]), так и Канадский национальный популяционный опрос состояния здоровья (Canadian National Population Health Survey [32]) используют HUI3.
3. Опрос по результатам лечения, проводимый Центрами Medicare и Medicaid (US Centers for Medicare and Medicaid Services Health Outcomes Survey [33]), ранее использовал SF-36 (первую версию), а теперь применяют разработанный Управлением по делам ветеранов SF-12, который может позволить рассчитать оценки шкалы SF-6D, а также может использоваться для расчета оценок шкалы EQ-5D.
4. В Великобритании за последние 15 лет, правительственные агентства периодически проводили популяционные опросы, с применением EQ-5D для того, чтобы дать основания для принятия решений, кроме того, индивидуальные исследователи также проводили исследования на национальном уровне по оценке состояния здоровья населения, в качестве примера можно привести применение EQ-5D в работе *Kind* [34].
5. В США, поддерживаемый федеральным правительством проект по созданию американских весовых коэффициентов для EQ-5D привел к проведению нескольких популяционных исследований, которые сравнивали полученные суммарные оценки предпочтений с таковыми, полученными с помощью EQ-5D и британскими весовыми коэффициентами, а также со оценками, полученными на основе HUI2 и HUI3 [35].
6. *Fryback и соавт.* [7] опубликовали репрезентативные для США оценки предпочтений по всем шести измерительным системам (EQ-5D, SF-36, HUI2, HUI3, QWB (версия для самостоятельного

заполнения), и HALex), что являлось частью подержанного федеральным правительством исследования по созданию национальной системы измерения здоровья (National Health Measurement Study).

7. Сеть по развитию системы измерения показателей исходов, сообщаемых пациентом (Patient Reported Outcome Measurement Information System Network) включила EQ-5D в этот большой проект по созданию набора вопросов, которые позволили бы оценить боль, слабость, физическое функционирование, эмоциональный стресс и социальное функционирование. Недавно были разработаны предсказательные модели с использованием общих и частных вопросов, включенных в этот банк для оценки предпочтений различных состояний здоровья [36].
8. В том случае, если эти «описательные системы» состояний здоровья будут радикально изменены или отброшены в поисках полностью пересмотренной модели QALY или другого показателя, наша возможность сравнивать прошлые, настоящие и будущие исследования будет поставлена под угрозу или вообще уничтожена. С другой стороны, если будут проводиться постепенные изменения в этих системах с необходимым вниманием к поддержанию возможности сравнения, у нас сохранится возможность сравнивать полученные результаты во времени.

Имеются ли другие аспекты определения и описания состояний здоровья, которые требуют дополнительного внимания исследователей? Следующая проблема продолжает провоцировать дебаты:

Основные системы измерения состояний здоровья имеют значительные различия в структуре доменов, некоторые из которых, как кажется, приводят к тому, что получается «одно и то же», но разными путями. Это может иметь негативные последствия для сравнимости оценок состояний здоровья (и QALY), полученных с помощью различных измерительных систем.

В качестве примера укажем, что домены для четырех хорошо известных измерительных систем названы (их создателями) следующим образом: EQ-5D (Подвижность, Самообслуживание, Обычная активность, Боль/Дискомфорт, Тревога/Депрессия); HUI 3 (Зрение, Слух, Речь, Способность находиться вне стационара, Сноровка, Эмоции, Мышление, Боль); QWB (Подвижность, Физическая активность, Социальная активность, набор комплексов симптомы-проблемы); и SF-6D (Физическое функционирование, Ролевые ограничения, Социальное функционирование, Боль, Психическое здоровье, Энергия). Fryback и соавт. [7] подтверждают, что, когда эти измерительные системы применяются на одной и той же популяции, они приводят к значимо различным

оценкам состояния здоровья. Поэтому данные системы не дадут, в общем случае, одни и те же оценки QALY, и отношения стоимость/выгода, если даже будут применяться в одной и той же выборке.

С другой стороны, такое многообразие дает аналитику возможность выбрать инструмент, наилучшим образом подходящий для анализируемой проблемы в области здоровья. На самом деле сложно сформулировать *a priori*, что представляет собой «правильный» (т.е. необходимый и достаточный) набор изучаемых состояний здоровья в отсутствие некоего набора внешних критериев, таких как желание сопоставить структуру инструмента с доменами здоровья, которые с наибольшей вероятностью будут меняться под влиянием вмешательства. Такие критерии также, вероятно, будут зависеть от целей и задач исследования, например, акцентируется ли оно на вмешательстве, направленном на специфическое заболевание или сравнивает разные вмешательства, направленные на разные заболевания.

В противоположность этому существуют хорошо определенные психометрические подходы для оценки и улучшения содержания вопросов внутри выбранного домена — хорошая новость, особенно в свете следующей проблемы.

У инструментов для измерения здоровья имеется тенденция к появлению эффектов «потолка и пола» (ограниченности значений) и, в целом, полагаться на набор вопросов, которых слишком мало, для того, чтобы обеспечить адекватное отражение всего многообразия исходов в каждом из доменов здоровья. Сила выраженности этой проблемы имеет непосредственные последствия для оценки QALY

Проблема адекватного количества и содержания вопросов была признана ряд разработчиков инструментов. Например, изменения, произошедшие в SF-36 при переходе от первой ко второй версии [37] включали в себя, например, усиление наполненности некоторых вопросов (переход к выбору из пяти вариантов ответов вместо двух для семи вопросов двух шкал ролевого функционирования) и упрощение содержания в других (переход от шести вариантов ответов к пяти в наборах вопросов для шкал психического здоровья и энергичности). HUI3 был разработан так, чтобы решить проблемы, имевшиеся у HUI2, при этом были изменены, как вопросы, так и суммарные шкалы, чтобы улучшить применение инструмента в клинических и популяционных исследованиях и усилить структурную независимость доменов (что улучшает надежность алгоритмов расчёта суммарной оценки) [38]. При использовании обоих вариантов инструмента у больных сахарным диабетом, Maddigan и соавт. [39] обнаружили, что HUI3 дает больший диапазон возможных оценок состояния здоровья и лучше разделял пациентов в соответствии с клинически выраженной степенью имеющихся нарушений.

Недавно, *Pickard et соавт.* [40] представили прекрасный пример того, как использование современных психометрических техник позволяет оценить, насколько постепенные изменения, вносимые в структуру инструмента для измерения состояния здоровья, улучшают измерительные свойства самого инструмента. Они изучили эффект от перехода со стандартного, трехуровневого формата EQ-5D на пятиуровневый формат, по каждому из пяти доменов, использующихся в этом инструменте. Используя модель латентных признаков Раша (один из вариантов теории шкалирования) на американской и голландской выборках *Pickard и соавт.* сумели установить, какие уровни в инструментах 5-L и 3-L являются статистически эквивалентными (таким образом, обеспечив связь между инструментами), а также подтвердив, что 5-L расширяет покрытие анализируемого спектра состояний здоровья для каждого домена. Успешные постепенные изменения, вносимые в SF-36, HUI, и EQ-5D (продолжающиеся) позволяют предположить, что компонент QALY, связанный с измерением состояния здоровья может эволюционировать так, чтобы улучшить научные свойства инструмента, не прерывая связи с оригинальным инструментом. Аналогичным образом, поддержание связи на основе моделей, пришедших из теории шкалирования и других техник, позволяет новым и старым показателям измеряться в одних единицах.

Особенно многообещающим в этом процессе улучшения является использование моделей, базирующихся на теории шкалирования. Их использование *Pickard et соавт.* [40] показывает, чего можно добиться, подвергая жесткому психометрическому анализу на основе современных измерительных подходов, вносимые в вопросы изменения. Для каждой шкалы в многомерном наборе шкал, который и представляет собой систему измерения состояния здоровья, можно совместно тестировать на наличие ограничительных эффектов («пола и потолка»), уточнять, достаточно ли в системе вопросов по количеству и содержанию для адекватного «покрытия» исследуемого домена здоровья, и насколько шкала сама по себе является (достаточно) одномерной, как должно быть в измерительной системе. Строго говоря, домен-специфические шкалы для каждой из использующихся сейчас измерительных систем должны пройти такое исследование.

В будущем, такой анализ на основе теории шкалирования, может открыть путь для создания новых, более адекватных наборов вопросов для каждого домена здоровья, аналогично тому, как это было сделано для методов оценки состояния здоровья не базирующихся на предпочтениях в поддержанном Национальными институтами здоровья США проекте по разработке системы измерения показателей исходов, сообщаемых пациентом (PROMIS) [41]. После того, как набор вопросов для каждого домена

был откалиброван, можно переходить к получению набора репрезентативных оценок предпочтений для каждого вопроса, шкал и, в конце концов, композитного индекса состояния здоровья с использованием стандартных методов оценки предпочтений. *Revicki и соавт.* [36] вывели уравнение, которое может базироваться на глобальных и специфических наборах вопросов системы измерения показателей исходов, сообщаемых пациентом для оценки показателей шкалы EQ-5D. Другой пример применения теории шкалирования для перехода от результатов одного инструмента к другому обсуждается в следующем разделе.

Оценка состояний здоровья

Каждая из основных систем измерения состояний здоровья располагает процедурой шкалирования, благодаря которой использованная в инструменте многомерная модель состояния здоровья превращается скалярную суммарную оценку, фактически $V(H_{st})$ в уравнении традиционного QALY. Но методика расчета суммарной оценки различается между системами и это потенциально может оказаться очень важным. Более того, некоторые аспекты методологии оценки состояний здоровья, общие для всех систем были поставлены под сомнение, как отмечалось в первом разделе.

С нашей точки зрения, эти вопросы по оценке состояния здоровья могут быть наиболее эффективно решены путем использования традиционного QALY как стартовой точки для дальнейшего анализа. Разведывательный анализ может затем быть проведен для того, чтобы (1) лучше понимать и нивелировать различия между существующими системами измерения здоровья или (2) добиваться постепенного — а может быть, и более, чем постепенного — улучшения QALY, сохраняя способность учитывать огромную литературу по использованию QALY. В определенной степени мы получаем очень ценную перспективу с точки зрения влияния любых изменений в процедуре оценки QALY на такие действительно значимые показатели, как отношения стоимость/полезность и оценку состояния здоровья населения.

Распознавание и корректировка различий в модели традиционного QALY

Проблемы с оценкой $V(H_{st})$ в традиционной модели QALY которые заслуживают дальнейшего анализа связаны как с измерением, так и с агрегированием весов в каждой из используемых систем и изучением, и возможной коррекцией различий суммарных оценок, полученных с помощью различных систем. Мы сейчас кратко остановимся на этих вопросах, понимая, что каждый из них заслуживает обсуждения в рамках отдельной статьи.

Альтернативные методы изучения предпочтений. Оценки для QWB были получены на основе процедуры ранжирования [11]; для EQ-5D на основе процедуры временной эквивалентности (time-trade-off (ТТО)) [10]; для HUI2/3 с использованием как процедуры стандартной азартной игры (standard gamble (SG)), так визуальной аналоговой шкалы (VAS) [9]; а для SF-6D использовалась процедура стандартной азартной игры [12]. Временной горизонт, использующийся в вопросах, направленных на оценку предпочтений также варьирует в значительных пределах, так, например, респондентов, опрошенных при создании QWB просили представить ощущения в каждом из состояний здоровья на протяжении одного дня, тогда как респонденты, которых опрашивали при создании EQ-5D (как в США, так и в Великобритании) исходили из временного горизонта в 10 лет, когда они проводили оценку состояний с помощью метода временной эквивалентности. Все измерительные системы полагались при расчете весовых коэффициентов предпочтений на выборку респондентов из общей популяции (в противоположность, например, субпопуляциям с определенным заболеванием или инвалидностью); но сами популяции и периоды сбора информации различались весьма значительно [1, 9–12].

Альтернативные подходы к расчету суммарной оценки для (многомерного) состояния здоровья. В HUI2/3 суммарная (точечная) оценка состояния здоровья для индивидуума в определенной позиции по каждой из шкал здоровья были получены с помощью моделирования на основе теории многопараметрической полезности (multi-attribute utility theory, MAUT) используя как мультипликативную, так мультилинейную формы. Для EQ-5D (как варианта, базировавшегося на американских, так и для варианта, основанного на британских данных), QWB и SF-6D суммарные оценки были получены путем эконометрического моделирования: полученные на выборке оценки состояний здоровья были подвергнуты регрессионному анализу, в котором независимыми переменными были уровни здоровья (атрибуты) и, таким образом, была создана модель, предсказывавшая значение суммарной оценки на основе любой комбинации уровней здоровья. *Petrillo и Cairns* [42] приводят полезный обзор методологических вопросов, возникающих при использовании различных подходов к получению суммарной оценки состояний здоровья.

Состояния, хуже смерти. Чрезвычайно важно, что основные системы различаются по тому, признают ли они существование состояний, которые хуже смерти (или, точнее быть мертвым) и, если да, то как они их оценивают. Как EQ-5D, так и HUI2/3 сконструированы таким образом, чтобы разрешить существование таких состояний, которые имеют отрицательное значение ценности (при этом смерть

продолжает иметь нулевое значение) в то время как ни QWB, ни SF-6D (а также HALex) такого не позволяют. Очевидно, что тут возникают сложные философские и моральные проблемы. Даже с более прагматичной точки зрения основные измерительные системы могут прийти к абсолютно разным выводам, с точки зрения общества, по поводу состояния здоровья лиц, находящихся в особенно серьезных или тяжелых клинических состояниях.

Существующие различия между системами оценки состояния здоровья в присвоении весовых оценок в сочетании с различиями в определении состояний здоровья и конструкции инструментов, отмеченные выше означают, что они, в общем случае, будут давать различные суммарные оценки (различные значения QALY) в каждом конкретном случае, будь то анализ стоимость-полезность или оценка здоровья населения.

Из сложившейся ситуации имеется, как минимум, три выхода.

Во-первых, можно признать, что наличие различных инструментов является положительным фактом, который проливает дополнительный свет на действительно сложную проблему оценки лет жизни, откорректированных на качество. При этом всегда имеется возможность выполнить анализ чувствительности и определить, насколько сильно влияет замена одной измерительной системы на другую на расчеты показателей стоимости-полезности или индикаторов популяционного здоровья. Некоторые недавние примеры (из большого числа возможных работ) включают исследование *Franks и соавт.* [43] о влиянии выбранной системы измерения на расчет показателей стоимости-эффективности для одной проблемы со здоровьем и для разных проблем; работу *Fryback и соавт.* [7] анализировавших, как рассчитанный популяционный уровень здоровья различается при использовании разных измерительных систем; работу *Janssen и соавт.* [44] сравнивших эффективность EQ-5D и HUI2/3 в одной и той же выборке на основе показателей теории информации и работу *Stevens и соавт.* [45] исследовавших различия в предиктивной валидности HUI2 в зависимости от того, как производился расчет суммарной оценки — на основе многомерной теории полезности (MAUT) или регрессионного подхода.

Во-вторых, можно попробовать перекодировать суммарные оценки, полученные в одной системе в другую или (более широкомасштабно) из каждой системы во все другие. Строго говоря, это означает пересчет $V(Hst)$ из одной системы (например, SF-6D) в другую систему (например, EQ-5D). Если это возможно сделать, то различные показатели QALY, получающиеся в результате применения различных измерительных систем могут быть «сопоставлены» и представлены в аналогичных единицах или, по крайней мере, напрямую сравнимы на одной и той

же выборке респондентов. Перекодировка может оказаться практичным, конструктивным ответом на современную реальность многочисленных, конкурирующих друг с другом систем измерения состояния здоровья и некоторые недавно опубликованные статьи, и доклады на конференциях показывают, как это может происходить.

Возможны два основных направления при перекодировке суммарных шкал. Самым очевидным является использование одной или более «обучающих выборок», которые будут использоваться для построения статистической модели (с использованием корреляционного или регрессионного анализа) взаимосвязи между суммарными оценками для любой пары имеющихся инструментов. В идеале предсказательную валидность этих моделей необходимо будет проверить на других выборках. Используя данные Национального исследования состояния здоровья (National Health Measurement Study), *Fryback и соавт.* [46] продемонстрировали возможность такого подхода, создав предиктивные модели, которые позволяют проводить попарное сравнение пяти индексов, основанных на предпочтениях: EQ-5D, HUI2, HUI3, QWB-SA и SF-6D. Необходимо также отметить, что опубликованные данные, устанавливающие связь между оценками шкал SF-12 и EQ-5D [28—30] могут позволить аналогичную (хотя и двухступенчатую) перекодировку, поскольку оценки SF-6D могут быть напрямую рассчитаны из SF-12. Второй многообещающий подход к перекодировке разработанный и первоначально использованный *Fryback и соавт.* [46], использует иерархическое моделирование в рамках теории шкалирования, когда в модель включаются одновременно все пять индексов и устанавливаются статистические связи, позволяющие провести попарную перекодировку результатов. В рамках этого подхода, значение суммарной оценки по одной шкале (например, HUI3) сопоставляется с латентной переменной, которая отражает многомерную шкалу здоровья (как принято в теории шкалирования), и затем предсказанное значение другой шкалы (например, SF-6D) может быть получено на основе этого значения переменной.

Третьим возможным ответом на многочисленность инструментов для определения состояния здоровья, используемых при расчете традиционного QALY является инициация процесса согласования, который позволит идентифицировать среди существующих «кандидатных» систем измерения состояния здоровья «референтную» (которая будет использоваться для расчета «референтного» QALY). Такой подход будет аналогичен подходу, использованному американской Комиссией по стоимости-эффективности в медицине и здравоохранении (US Panel on Cost-Effectiveness in Health and Medicine) которая установила стандартный набор методологических подходов в анализе стоимости-эффективности, который по-

зволяет улучшить техническое качество и сопоставимость исследований [2]. В рамки данной статьи не входит анализ преимуществ, стоимости и возможности осуществления такой стратегии, однако любые рассуждения на данную тему должны базироваться на всеобъемлющей, беспристрастной оценке систем измерения, причем критерия для оценки должны быть сформулированы заранее. Возможные критерии, которые были успешно использованы при оценке не использующих предпочтений инструментов измерения QALY в онкологии и других областях включают те, что были опубликованы Фондом изучения медицинских исходов (Medical Outcomes Trust) [47, 48].

Другие проблемы оценки состояний здоровья

Значительная часть обсуждения и критики традиционной модели QALY вращается вокруг вопросов оценки состояний здоровья, в наиболее широком понимании, и мы хотим осветить еще две проблемы.

Во-первых, даже сторонники традиционной модели QALY признают, что имеется ряд упрощающих допущений, которые встроены в модель и которые, как минимум, требуют дальнейшего анализа. Например, предполагается, что ценность пребывания в некотором состоянии на протяжении двух лет в два раза больше, чем в течение одного года (иногда называется эффектом постоянства количества) — если только не применяется экспоненциальный фактор дисконтирования, который используется для приведения получаемых значений к настоящему времени. Один часто обсуждаемый, формальный ответ на озабоченность тем фактом, что расчет традиционного QALY может недостаточно адекватно отражать предпочтения на индивидуальном уровне — это подход с использованием эквивалента здоровых лет жизни (см. *Mehrez и Gafni* [14]).

Недавно *Salamon и Murray* [49] разработали и применили подход с использованием нескольких методов для того, что совместно проанализировать и сравнить основные методы изучения предпочтений (стандартную азартную игру, метод временной эквивалентности, визуальную аналоговую шкалу и метод эквивалентности жизней (person trade-off (PTO)) с точки зрения их способности позволить рассчитать весовые коэффициенты для состояний здоровья с учетом таких показателей, как отношение к риску, временные предпочтения и проблему распределения ресурсов

В целом, дополнительные эмпирические данные необходимы по поводу того, насколько хорошо индивидуальные предпочтения разных профилей здоровья описываются откорректированными на предпочтения по времени суммами предпочтений отдельных компонентов здоровья, как это делается в традиционном QALY. Несмотря на технические и интеллектуальные препоны, которые стоят перед такой всеобъемлющей оценкой исходов воздействий

на здоровье, выигрыш будет значительным. Экспериментальный анализ может пролить свет на разумность допущений об изменении предпочтений во времени и предпочтения различных состояний здоровья, которые встроены в традиционное QALY, например, упоминавшийся выше эффект постоянства количества, который предполагает, что дополнительное время пребывания в некоем состоянии здоровья не зависит от предшествующей длительности пребывания в этом состоянии.

Для того, чтобы всеобъемлющая оценка таких профилей могла быть сравнена с оценкой, лежащей в основе традиционного QALY, состояния здоровья, составляющие основу каждого профиля, могут быть взяты из наборов состояний, которые представлены в современных системах оценки состояния здоровья (т.е. из HUI, EQ-5D, QWB). Адекватное сравнение всеобъемлющего подхода и традиционного QALY требует, чтобы профили сравнивались в одном и том же временном промежутке. Как подобное сравнение можно выполнить на практике, продемонстрировано в работе *Lipscomb* [50], в которой использовались состояния здоровья, взятые из QWB.

Вторым важным критическим замечанием является то, что все ранее обсуждавшиеся подходы предполагают, что ценностный компонент модели должен основываться на предпочтениях, основные различия заключаются в том, чьи это должны быть предпочтения и как они оцениваются. Исследователи, работающие с традиционной моделью QALY, обычно отмечают, что $V(H_{st})$ должно быть получено на популяционной выборке, отражая заранее сформулированные предпочтения репрезентативной выборки из общей популяции. Однако для проекта «эквивалентов сохраненных молодых жизней» (SAVEs), *Nord* предлагал, чтобы предпочтения были бы сформулированы лицами, которые находились в том состоянии здоровья, который оценивается (не просто случайная популяционная выборка).

Тем не менее, *Dolan u Kahneman* [16], а также *Hausman* [17] поставили под сомнение идею о том, что индивидуальные данные должны базироваться на предпочтениях. *Dolan u Kahneman* утверждают, что предпочтения, полученные до возникновения события или после имеют тенденцию давать смещенные результаты оценки ценности, относительно той, которую индивидуум присваивает состоянию здоровья в тот момент, когда он его испытывает. С их точки зрения адекватной аналитической задачей является оценить ценность именно в тот момент, когда человек испытывает некое событие, а не в тот, когда он ожидает его наступления или вспоминает о нем (отсюда, они подчеркивают важность использования «полезности опыта» в противоположность «полезности принятия решения»). С другой стороны, *Hausman* не признает ни один из этих подходов к оценке полезности здоровья с точки зрения его участия в фор-

мировании благополучия человека. Вместо этого он призывает к разработке стратегий распределения ресурсов здравоохранения, которые бы оптимизировали возможности индивидов вести продуктивную и достойную жизнь. В целом, здоровье у него рассматривается как промежуточный компонент обеспечивающий возможность индивидуума осуществлять важные жизненные проекты, тогда как QALY просто оценивает связанное со здоровьем благополучие.

По причине наличия этих и ряда других важных критических возражений в отношении традиционного QALY, ясная и своевременная задача заключается в разработке альтернативных моделей, которые, в конце концов, смогут помочь в принятии решений обществом более полезным и менее дискуссионным способом, нежели традиционная модель.

В то же время, работа над разрешением описанных выше проблем должна проходить таким образом, чтобы модифицированные модели QALY — вне зависимости от формы, которую они примут — могли бы быть связаны с традиционной моделью. Это будет способствовать проведению анализа чувствительности, который позволит установить, насколько предпринятые изменения в оценке состояний здоровья влияют на создание показателей здоровья для анализа стоимости-эффективности, мультиплицирования популяционного здоровья и других практических задач. Таким образом мы будем находиться в ситуации, когда сможем оценить, к каким изменениям приводят изменения в QALY.

Необходимость принимать во внимание вопросы равенства, справедливости и адекватности распределения ресурсов

Никто не возражает против мнения о том, что вопросы равенства, справедливости и адекватного распределения ресурсов играют роль (или должны играть роль) при принятии решений в области здравоохранения и что традиционная модель QALY не позволяет учитывать напрямую эти показатели при анализе стоимости-эффективности или других видах экономического анализа. Тем не менее, имеются различные взгляды на тему того, каким наилучшим образом аналитически подойти к данной проблеме и имеется как минимум два возможных пути вперед.

Одним подходом является модификация весовой схемы, используемой в модели QALY таким образом, чтобы весовые коэффициенты ценности варьировали в соответствии с характеристиками тех лиц, кто находится в тех или иных состояниях здоровья, а не отражали усредненные, популяционные оценки здоровья. Таким образом $V(H_{st})$ модели традиционного QALY превращается в $V(H_{stx})$, где x обозначает определенные характеристики конкретного индивида. В этой связи *Nord* проанализировал проблему допущения традиционного QALY, что общественная

значимость исхода вмешательства в области здравоохранения для индивида пропорциональная выигрышу ценности для этого индивида (традиционное DQALY) — вне зависимости от исходной тяжести состояния, возраста и других факторов. В качестве альтернативы он предложил использовать метод эквивалентности человеческих жизней (РТО), для того, чтобы получить ценности, которые принимают во внимание указанные выше параметры и использовать показатели SAVE в качестве альтернативной «валюты» при анализе [15, 51]. *Ubel и соавт.* [16] обсуждают подходы к оценке стоимости-эффективности с помощью напрямую полученных общественных ценностях, а не на QALY

Альтернативным подходом для того, чтобы учитывать проблему справедливости при принятии решения о распределении ресурсов является иерархический подход: дополнительные требования по равному распределению (или ограничения) применяются к традиционной модели QALY на втором, четко определенном этапе, а не вкраплены в весовые коэффициенты предпочтений модели. Это позволяет определить воздействие на эффект (общее количество выигранных QALY) и равенство (их распределение в изучаемой популяции) с учетом принимаемых решений о распределении ресурсов. Подобный подход позволяет оценить взаимоотношения между эффективностью и справедливостью, выяснить, что теряется, когда применяются те или иные правила равенства или справедливого распределения. Исходные данные для оценки важности вопросов равенства могут быть получены с помощью изучения мнения населения, как описано в конце данного раздела.

Мы считаем, что второй подход является более разумным, поскольку его можно представить, как естественное дополнение или улучшение модели традиционного QALY. В рамках этого подхода можно оценить выигрыш в равенстве, а также подсчитать возможные потери эффективности (чистые QALY), при переходе от традиционной модели QALY к альтернативной форме, которая учитывает вопросы равенства распределения ресурсов. Соответственно, могут быть оценены альтернативные издержки (в виде потерянного улучшения здоровья) большей справедливости.

Как такой иерархический процесс, основывающийся на модели традиционного QALY может быть сформулирован аналитически? Существуют по крайней мере две широких стратегии.

Весовые коэффициенты, учитывающие равенство

Среди нескольких публикаций на эту тему выделяется недавняя работа *Bleichrodt и соавт.* [52] в которой была предпринята попытка разработать «ранговую» модель QALY, которая позволяет при-

менить весовые коэффициенты равенства к любому возможному профилю QALY, который может наблюдаться у N человек в популяции. Этим профилем является упорядоченный вектор, отсортированный в порядке убывания значений традиционных QALY для этих N индивидов, в зависимости от изучаемого вмешательства и других допущений.

Таким образом, связанные со здоровьем исходы, которые, как ожидается, будут являться следствием двух конкурирующих вмешательств, могут быть сравнены с точки зрения общества путем сравнения откорректированных на равенство QALY. Особыми случаями и потенциальными вариантами этого подхода является «QALY утилитаризм» (то, что получается, когда в анализе стоимости-эффективности используется традиционная модель QALY), использование функции социального благополучия по Роулсу, в которой все весовые коэффициенты равенства присваиваются индивидам, которым хуже всего и подход к установлению приоритетов по «чувству внутренней справедливости» Вильямса.

Wagstaff предложил измерять степень неприятия обществом неравенства в здоровье путем использования отношения, которое будет показывать, от какого количества традиционных QALY общество готово отказаться для достижения определенного выигрыша в равенстве [53].

В ответ на утверждения некоторых, что обычная система максимизации QALY приводит к дискриминации инвалидов и хронических больных, *Johannesson* [54] предложил альтернативную форму анализа стоимости-эффективности, в котором максимизируются относительные, а не абсолютные изменения QALY. Параметр «относительных изменений» для группы пациентов определенного возраста и пола рассчитывается как усредненное ожидаемое число QALY для этой группы деленное на усредненное ожидаемое число QALY в исходном состоянии. Две группы пациентов с одинаковыми параметрами относительных изменений будут иметь одинаковые весовые коэффициенты равенства при проведении анализа стоимости-эффективности вне зависимости от абсолютного числа QALY, которое может быть выиграно в результате вмешательства.

Необходимо также отметить использование *Nord и соавт.* [51] многоступенчатого подхода с применением весовых коэффициентов равенства (точнее, весовых коэффициентов, отражающих относительную тяжесть заболевания и относительный потенциал для улучшения здоровья), которые были получены на выборке из общей популяции с применением техники эквивалентности жизней (РТО). При этом весовые коэффициенты для состояний здоровья в этом «анализе стоимость-ценность» (см. также [15]) были получены при помощи методики временной эквивалентности у лиц, которые бывали в этих состояниях здоровья, а не от общей популяции. Однако ввиду

таких особенностей дизайна, возможно, и специально, этот анализ не дает ясную связь с традиционной моделью QALY.

Модели оптимизации в условиях ограничений

Хорошо известно, что любой анализ стоимости-эффективности может быть переформулирован как задача математического программирования в которой необходимо выбрать оптимальный набор вмешательств, который максимизирует улучшение состояния здоровья в условиях бюджетных ограничений. Если объективная функция описывается в понятиях традиционного QALY все сводится к проблеме линейного (или целочисленного) программирования, которая четко соответствует стандартной модели стоимости-полезности. На протяжении ряда лет стало понятно, что возможно также что на этапе оптимизации можно внести дополнительные ограничения, так, чтобы отразить вопросы равенства и адекватного распределения ресурсов. Например, если субпопуляция В исходно имеет суммарный показатель QALY меньше X, необходимо в этой популяции провести вмешательство Y вне зависимости от ожидаемого выигрыша в QALY. На самом деле, практически любой набор правил, относящихся к равенству или справедливости может быть переведен на алгебраический язык ограничений модели (и можно даже утверждать, что, если правило равенства и справедливости не может быть однозначно переведено на алгебраический язык, возможно, само правило не однозначно).

Исследования, которые изучали возможности и последствия внедрения такого подхода в аналитическую систему, применяемую для анализа стоимости-эффективности, включают работы *Epstein и соавт.* [55], *Stinnett и Paltiel* [56], а также *Chen и Bush* [57]. Во всех этих работах традиционная модель QALY фактически сохранялась, поэтому можно было рассчитать различия между максимальным приростом QALY, когда ограничения равенства и справедливости не применялись и максимальным приростом в условиях специфических ограничений. Такой подход кажется альтернативным и реализуемым подходом для определения границы равенство-эффективность по *Wagstaff* [53].

Менее формализованный и более интерактивный подход для идентификации проблем, связанных с равенством и справедливостью может быть предпринят путем консультаций с населением, когда обычные отношения стоимости-полезности изучаются представителями общественности на предмет их соответствия общественным предпочтениям. Такой процесс может свести вместе обычных людей (просто граждан или специально отобранных представителей общественности) с теми, кто предоставит им необходимые данные о программах в

области здоровья и результатах вмешательств, будут способствовать диалогу и помогут составить рекомендации, которые, по своей природе, могут быть как качественными, так и количественными. Одним из примеров такого подхода на национальном уровне является Гражданский Совет при национальном институте клинического совершенства Великобритании (UK NICE Citizens Council), состоящая из 30 человек группа, отражающая широкие слои населения. Решения Совета, обращающие внимание на факторы, которые должны влиять на распределение выигранных QALY в национальной системе здравоохранения Великобритании, учитываются комитетами NICE при принятии решений об оценке медицинских технологий [58, 59].

Заключительные положения

Основная тема данной работы — как сохранить и улучшить традиционную модель QALY — может рассматриваться как путь к достижению более важной цели. Эта цель заключается в улучшении способности измерять и оценивать здоровье популяций и отдельных лиц, а также оценивать влияние на здоровье различных вмешательств.

Мы обсудили, с разной степенью детализации, научную повестку дня по усилению научной обоснованности и полезности подхода с использованием QALY путём улучшения 1) систем, использующихся для описания состояний здоровья; 2) методов оценки здоровья и 3) нашей способности принимать во внимание вопросы равенства, справедливости и адекватности распределения ресурсов. На этом пути имеются важные научные и политические основания для того, чтобы относиться к традиционной модели QALY как к стартовой точке для развития и тестирования альтернативных моделей, основанных на предпочтениях — вне зависимости от того, являются ли предлагаемые изменения в традиционную модель QALY постепенными или резкими. Идя по этой дороге, мы учитываем тот опыт, который был приобретён за многие годы и отражен в многочисленных исследованиях, посвященных науке и искусству измерения и оценивания здоровья. Это также позволяет поддерживать непрерывность и способствовать сравнимости при отслеживании тенденций в изменении популяционного здоровья и при проведении анализа стоимости-эффективности для того, чтобы идентифицировать те вмешательства, которые предлагают хорошее соотношение цены и качества.

Очевидно, что имеются альтернативные подходы к оценке исходов вмешательств и приносимых ими выгод в отношении здоровья, включая анализ стоимости-выгоды (willingness-to-pay (WTP)) на основе методологии готовности заплатить и экспериментов по дискретному выбору (discrete choice experiments (DCE)) на основе совместного (конджойтного) ана-

лиза, который позволяет ранжировать вмешательства. По мере того, как исследования и эксперименты в этих областях продолжаются, исследователи не только будут сталкиваться со своими собственными методологическими и эмпирическими проблемами, но могут пожелать (или должны захотеть) иметь показатели для сравнения. Традиционная модель QALY является одним из таких естественных источников сравнения для того, чтобы определить, например, насколько, в данной задаче распределения ресурсов,

базирующийся на согласии заплатить анализ стоимость-выгода, оценка на основе эксперимента по дискретному выбору или базирующийся на QALY анализ стоимости-эффективности дают сходные или сильно различающиеся рекомендации.

По многочисленным причинам мы должны сохранить QALY, улучшить эту модель и более широко подходить к улучшению научной обоснованности и полезности базирующихся на предпочтениях подходов к измерению здоровья.

Литература

1. Drummond M., Sculpher T.G., O'Brien B.S. *Methods for the Economic Evaluation of Health Care Programmes* (3rd ed.). New York: Oxford University Press, 2005.
2. Gold M., Siegel R., Weinstein M. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
3. National Institute for Health and Clinical Excellence (NICE). Available from: <http://www.nice.org.uk> [Accessed January 24, 2009].
4. O'Donnell J., Pham S., Pashos C., Miller D. Health technology assessment: Lessons learned from around the world. *Value Health* 2009;12(Suppl.) в печати.
5. ISPOR. ISPOR Thirteenth Annual International Meeting Abstracts. *Value Health* 2008;3:A1—311.
6. Statistics Canada. Canadian Community Health Survey. September 2000 and Ongoing. Available from: <http://www.statcan.gc.ca> [Accessed January 24, 2009].
7. Fryback D.G., Dunham N.C., Palta M., et al. Norms for six generic health-related quality-of-life indexes from the national health measurement study. *Med Care* 2007;45:1162—70.
8. Fryback D.G., Lawrence W.F., Martin P.A., et al. Predicting quality of well-being scores from the SF-36: results from the Beaver Dam Health Outcomes Study. *Med Decis Making* 1997;17:1—9.
9. Feeny D., Furlong W., Torrance G.W., et al. Multiattribute and single-attribute utility functions for the Health Utilities Index Mark 3 System. *Med Care* 2002;40:113—28.
10. Brooks R., Rabin R., de Charro F. *The Measurement and Valuation of Health Status Using EQ-5D: a European Perspective*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.
11. Andresen E.M., Rothenberg B.M., Kaplan R.M. Performance of a self-administered mail version of the Quality of Well-Being (QWB-SA) Questionnaire among older adults. *Med Care* 1998; 36:1349—60.
12. Brazier J.E., Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42:851—9.
13. Erickson P. Evaluation of a population-based measure of quality of life: the Health and Activity Limitations Index (HALex). *Qual Life Res* 1998;7:101—14.
14. Mehrez A., Gafni A. Quality-adjusted life years, utility theory, and healthy years equivalents. *Med Decis Making* 1989;9:142—9.
15. Nord E. *Cost-Value Analysis in Health Care: Making Sense out of QALYs*. New York: Oxford University Press, 1999.
16. Ubel P., Nord E., Gold M., et al. Improving value measurement in cost-effectiveness analysis. *Med Care* 2000;38:892—901.
17. Dolan P., Kahneman D. Interpretations of utility and their implications for the valuation of health. *Econ J* 2008;118:215—34.
18. Hausman D.M. Valuing health properly. *Health Econ Policy Law* 2008;3:79—83.
19. Miller W., Robinson L.A., Lawrence R.S., eds. *Valuing Health for Regulatory Cost-Effectiveness Analysis*. Washington DC: Institute of Medicine, National Academy Press, 2006.
20. Neumann P.J. *Using Cost-Effectiveness Analysis to Improve Health Care*. New York: Oxford University Press, 2005.
21. US Food and Drug Administration. Guidance for industry— patient-reported outcome measures: use in medical product development to support labeling claims. 2006. Available from: <http://www.fda.gov/cber/gdlns/probl.pdf> [Accessed January 24, 2009].
22. U.S. Federal Registry. Medicare program: criteria and procedures for extending coverage decisions that relate to health care technology. *US Federal Register* 1989;54:4302—18.
23. Feeny D. The roles of preference-based measures in support of cancer research and policy. In: Lipscomb J., Gotay C.C., Snyder C., eds. *Outcomes Assessment in Cancer: Measures, Methods, Applications*. Cambridge: Cambridge University Press, 2005.
24. Brauer C.A., Rosen A.B., Greenberg D., Neumann P.J. Trends in the measurement of health utilities in published cost-utility analyses. *Value Health* 2006; 9:213—8.
25. Szende A., Williams A., eds. *Measuring Self-Reported Population Health: an International Perspective Based on the EQ-5D*. Budapest: Spring Med Publishing, 2004.
26. Bryan S., Sofaer S., Siegelberg T., Gold M.R. Has the time come for CEA in U.S. Health Care? *J Health Econ Policy Law* in press.
27. US Agency for Healthcare Research and Quality. Medical Expenditure Panel Survey (MEPS). Available from: <http://www.meps.ahrq.gov> [Accessed January 24, 2009].
28. Sullivan P.W., Ghushchyan V. Mapping the EQ-5D Index from the SF-12: U.S. general population preferences in a nationally representative sample. *Med Decis Making* 2006;26:401—9.
29. Lawrence W.F., Fleishman J.A. Predicting EuroQol EQ-5D preference scores from the SF-12 Health Survey in a nationally representative sample. *Med Decis Making* 2004;24:160—9.
30. Franks P., Lubetkin E.L., Gold M.R., et al. Mapping the SF-12 to the EuroQol EQ-5D Index in a national U.S. sample. *Med Decis Making* 2004;24:247—54.
31. Statistics Canada and the US Centers for Disease Control and Prevention. Joint Canada—United States Survey of Health (JCUSH). Available from: http://www.cdc.gov/nchs/about/major/nhis/Canada_US.htm [Accessed January 24, 2009].
32. Statistics Canada. National Population Health Survey. Available at: <http://www.statcan.gc.ca/bsolc/olc-cel/olc-cel?lang=eng&catno=82-618-M> [Accessed January 24, 2009].
33. US Centers for Medicare and Medicaid Services. Medicare Health Outcomes Survey. Available from: <http://www.cms.hhs.gov/hos/> [Accessed January 24, 2009].

34. Kind P, Dolan P, Gudex C, Williams A. Variations in population health: results from a United Kingdom national questionnaire survey. *BMJ* 1998; 316: 736–41.
35. Shaw J.W, Johnson J.A., Coons S.J. U.S. Valuation of the EQ-5D Health States: development and testing of the D1 valuation model. *Medical Care* 2005; 43: 203–20.
36. Revicki D.A., Kawata A., Harnam N., et al. Predicting EuroQol (EQ-5D) scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items and domain item banks in a United States representative sample. UnitedBio-Source Corp. working Paper, November 2008.
37. Ware JE Jr. SF-36® Health Survey Update. Available from: [http:// www.sf-36.org/tools/sf36.shtml](http://www.sf-36.org/tools/sf36.shtml) [Accessed January 24, 2009].
38. Horsman J, Furlong W, Feeny D, Torrance G. The Health Utilities Index (HUI): concepts, measurement properties, and applications. *Health Qual Life Outcomes* 2003;1:54–66.
39. Maddigan S.L., Feeny D.H., Johnson J.A., for the DOVE Investigators. A comparison of the Health Utilities Indices Mark 2 and Mark 3 in type 2 diabetes. *Med Decis Making* 2003;23:489– 501.
40. Pickard A., Kohlmann T., Janssen M., et al. Evaluating equivalency between response systems: application of the Rasch Model to 3-Level and 5-Level EQ-5D. *Med Care* 2007;45:812–9.
41. National Institutes of Health. Patient-reported outcomes measurement information system: dynamic tools to measure health outcomes from the patient perspective. Available from: [http:// www.nihpromis.org](http://www.nihpromis.org) [Accessed January 24, 2009].
42. Petrillo J, Cairns J. Converting condition-specific measures into preference-based outcomes for use in economic evaluation. *Exp Rev Pharmacoeconom Res* 2008;8:453–6.
43. Franks P, Hanmer J, Fryback D.G. Relative disutilities of 47 risk factors and conditions assessed with seven preference-based health status measures in a national U.S. sample: toward consistency in cost-effectiveness analysis. *Med Care* 2006;44:478– 85.
44. Janssen M.F, Birnie E., Bonsel G.J. Evaluating the discriminatory power of EQ-5D, HUI2 and HUI3 in a U.S. general population survey using Shannon's indices. *Qual Life Res* 2007;16:895– 904.
45. Stevens K., McCabe C., Brazier J, Roberts J. Multi-attribute utility functions or statistical inference models: a comparison of health state valuation models using the HUI2 health state classification system. *J Health Econ* 2006;26:992–1002.
46. Fryback D, Palta M., Cherepanov D., et al. for the Health Measurement Research Group. Cross-walks among five self-reported summary health utility indexes: progress and prospects. Presented at the Annual Meeting of the Society for Medical Making, Pittsburgh, PA, Oct 24, 2007.
47. Scientific Advisory Committee for the Medical Outcomes Trust (Lohr K, et al.) Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res* 2002;11:193– 205.
48. Lipscomb J, Snyder C.F, Gotay C. Cancer outcomes measurement through the lens of the medical outcomes trust framework. *Qual Life Res* 2007;16:143–64.
49. Salomon J.A., Murray C.J. A multi-method approach to measuring health-state valuations. *Health Econ* 2006;13:281–90.
50. Lipscomb J. Time preference for health in cost-effectiveness analysis. *Med Care* 1989;27:S233–53.
51. Nord E., Pinto J.L., Richardson J., et al. Societal concerns for fairness in numerical valuations of health programmes. *Health Econ* 1999;8:25–39.
52. Bleichrodt H., Doctor J., Stolk E. A nonparametric elicitation of the equity-efficiency trade-off in cost-utility analysis. *J Health Econ* 2005;24:655–78.
53. Wagstaff A. QALYs and the equity-efficiency trade-off. *J Health Econ* 1991;10:21–41.
54. Johannesson M. Should we aggregate relative or absolute changes in QALYs? *Health Econ* 2001;10:573–7.
55. Epstein D.M., Chalabi Z., Claxton K., Sculpher M. Efficiency, equity, and budgetary priorities. *Med Decis Making* 2007;27: 128–37.
56. Stinnett A.A., Paltiel A.D. Mathematical programming for the efficient allocation of health care resources. *J Health Econ* 1996;15: 641–53
57. Chen M.M., Bush J.W. Maximizing health system output with political and administrative constraints using mathematical programming. *Inquiry* 1977;13:215–27.
58. NICE Citizens Council. NICE Citizens Council report: ultra orphan drugs. London, November 2004. Available from: [http:// www.nice.org.uk/](http://www.nice.org.uk/) [Accessed January 24, 2009]. [Alternatively or together, cite: Culyer AJ. NICE's Use of Cost-Effectiveness as an Exemplar of a Deliberative Process. *Health Econ Policy Law* 2006;1:299–318. and/or a critique of the process.]. 59 Gold MR, Sofaer S, Siegelberg T. Medicare and cost-effectiveness analysis: time to ask the taxpayer. *Health Aff (Millwood)* 2007;26:1399–406

Применение QALY в клинической практике и при принятии административных решений: проблемы и перспективы

Paul Kind,¹ Jennifer Elston Lafata, PhD,² Karl Matuszewski, MS, PharmD,³ Dennis Raisch, BSPHarm, MS, PhD⁴

¹ — University of York, York, UK

² — Center for Health Service Research, Henry Ford Health System, Detroit, MI, USA

³ — University HealthSystem Consortium, Oak Brook, IL, USA

⁴ — University of New Mexico, Albuquerque, NM, USA

Переводчик:

Виллом Ирина Александровна — ассистент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Колбин Алексей Сергеевич — д.м.н., профессор, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Ключевые слова: принятие решений на уровне пациента, откорректированная по качеству продолжительность жизни, экономический анализ

The Use of QALYs in Clinical and Patient Decision-Making: Issues and Prospects

Keywords: patient decision-making, QALYs, economic analysis

Введение

Лица принимающие решения (главные врачи, организаторы здравоохранения, менеджеры) используют результаты фармакоэкономического анализа на различных уровнях, включая медицинские учреждения, региональные и национальные министерства здравоохранения [1—7]. Прежде всего анализа полезности затрат (CUA — cost utility analysis). Краеугольным камнем этого анализа является количество лет жизни с поправкой на ее качество (QALY — quality-adjusted life-year), которое арифметически представляет собой произведение количества прожитых лет к ее качеству. Использование QALY как итоговой меры состояния здоровья удобно так же для инвесторов, менеджеров в системе здравоохранения и различных министерств, которые принимают решения с целью оптимизации затрат в условиях ограниченности финансовых ресурсов.

Тем не менее, важность и необходимость более широкого внедрения QALY в процесс принятия решений среди клиницистов и пациентов остается зачастую спорным. В рамках данной статьи принимают тот факт, что экономическая оценка медицинских технологий в большинстве стран мира является данностью. Следовательно, решающая роль фармакоэкономических исследований, как инструмента для принятия решений, вряд ли будет кардинально изменена в ближайшие 5—10 лет. Центральной проблемой в CUA является поиск итогового показателя. Возможно, это может быть QALY. Важнейшее значение для вычисления QALY имеет оценка влияния на заболеваемость, представленная посредством измерения HRQoL (health-related quality of life). Хотя эти данные используются в основном для экономической оценки новых технологий, они могут быть также использованы с другими целями, неэкономического характера, например, в мониторинге состоя-

ния здоровья отдельных пациентов, всего населения или отслеживании эффектов лечения в клинических исследованиях. В разных системах здравоохранения являются допустимыми измерения качества жизни при экономической оценке медицинских технологий. Некоторые из них предусматривают, что инструмент, используемый для оценки HRQoL, должен быть основан на социальных предпочтениях, отражающих взгляды населения в целом. Другие оперируют данными, полученными от пациентов или других лиц, непосредственно принимавших участие в медицинской технологии. Остается вопрос о том, чьи ценности должны быть учтены при разработке решений в обществе, в условиях данной системы здравоохранения. В конечном итоге, это вопрос политический. Он не определяется (и не должен определяться) экономистами здравоохранения или другими заинтересованными лицами.

В рамках настоящей статьи достаточно отметить, что адекватная оценка QALY может быть получена из нескольких источников, отражая взгляды как пациента, так и врачебного сообщества. Безусловно, существует некоторый диссонанс между политикой, стимулирующей развитие программ в сфере здравоохранения, и той помощью, которую реально предоставляют изо дня в день пациентам. Два данных противоречия объединяют посредством экономической оценки.

Решение врача определить пациента в больницу для лечения, применить инновационные методы терапии, выбрать направление в диагностики заболевания, по возможности, должны быть приняты максимально независимо от любой экономической оценки (при наличии её результатов). С другой стороны, программы в области здравоохранения, скорее всего, должны характеризоваться как эффективностью для здоровья, так и экономической целесообразностью предполагаемых вмешательств, при условии применения их в группах населения и в целых популяциях. В данной статье рассматриваются ряд связанных с этим вопросом, которые вытекают из более широкой интерпретации QALY, в которой признают его роль как составной мерой полезности для здоровья, но не полностью отражает или ограничена рамками рентабельности. При каких обстоятельствах определенная с помощью QALY ценность медицинской технологии для лиц принимающих решение будет соответствовать таковой ценности для врачей и других сотрудников здравоохранения в отношении пациентов? Является ли QALY в его нынешнем формате полезным для сравнения последствий для здоровья в контексте данной терапевтической проблемы? Как можно усовершенствовать получаемую с помощью QALY информацию для использования среди врачей, пациентов и при принятии решений первого уровня? И, наконец, какие препятствия необходимо преодолеть для распространения и по-

вышения эффективности QALY в неэкономических сферах разработки решений по медико-санитарной помощи?

QALY и разработка решений

Принимая QALY за составную часть экономической оценки и основной компонент CUA, остается открытым вопрос о том, имеет ли QALY значение для принятия решений в здравоохранении за пределами данной области. Вопрос может быть поставлен более широко: об общей роли экономической оценки в принятии решений в здравоохранении. Существует ли универсальный инструмент, применимый на всех уровнях принятия решений или же область использования каждого из показателей ограничивается определённой областью? Тогда целесообразно рассматривать QALY как сводный показатель полезности для здоровья, предназначенный, в первую очередь, для использования в данном контексте, но потенциально, и в других условиях. Для более широкого понимания самой полезности необходимо уяснить: 1). тип решений, принимаемых ответственными лицами по планированию, управлению, распределению ресурсов и оказанию медицинских услуг; 2). организационный уровень, на котором данные решения принимаются; 3). роль отдельных лиц и групп в данном процессе. На рис. 1 показано, как применение QALY может (с теоретической точки зрения) коснуться различных лиц, принимающих решения. В столбцах представлены различные уровни разработки решений в системе оказания медицинских услуг, а в строках — тип измерения / зона интереса.

Данная таблица помогает определить спектр потенциальных пользователей в различных секторах системы здравоохранения и ее управления, а также те уровни, на которых информация, полученная с помощью QALY, может быть использована для проведения различного рода наблюдений и измерений. Клетки, выделенные чёрным цветом, содержат примеры ситуаций, применение QALY в которых в настоящее время сомнительно, например, на правительственном уровне разработки решений относительно отдельных пациентов. В клетках без выделения приводятся примеры ситуаций, в которых применение QALY может иметь потенциал для практического принятия решений, например, в отношениях отдельных поставщиков и пациентов. Клетки с частичным выделением отражают неоднозначный потенциал QALY. Содержимое некоторых полей необходимо рассматривать только с теоретической точки зрения, так как в настоящее время не существует достаточной практики применения QALY в данных ситуациях. Например, для принятия решений в отношении конкретных пациентов, уместно рассматривать данные по продолжи-

тельности и качеству жизни конкретного пациента, а не QALY, полученный на основе совокупной оценки возможной пользы для здоровья всего общества. Набор областей применения QALY, показанный на рис. 1, должен быть уточнен в ходе дальнейшего изучения, здесь же его следует рассматривать в качестве иллюстрации.

Отметим, что в связи с ростом информированности населения и расширяющимся доступом к альтернативным возможностям, все большее распространение в литературе получает идея об участии пациента в процессе принятия решений в здравоохранении [8, 9]. Предметом исследований, целью которых явилось повышение эффективности данного участия с помощью HRQoL и показателей продолжительности жизни, было лечение онкологических заболеваний [3, 10–13], хронического гепатита С [14], сердечной недостаточности [15, 19], почечной недостаточности [16], болезни Паркинсона [17], серповидно-клеточная анемии [18]. Есть сообщения об использовании QALY при таких состояниях, как диабет [20] и онкологические заболевания [4–7]. Использование анализа полезности затрат, который подразумевает оценку исходов с помощью QALY, также распространяется и на фармацевтический бизнес [2–21].

Когда речь идёт об определении потенциальных сфер применения QALY, важны две основные характеристики модели. Прежде всего, с точки зрения клиницистов, в повседневной практике возможно некоторое снижение специфичности QALY. Во-вторых, система уравнивания, используемая для вычисления отношения количество / качество результата, наиболее часто калибруются на основании социальных предпочтений, а не индивидуальных взглядов пациента или других лиц, заинтересованных в лечении. Составные элементы в QALY, качество жизни и продолжительность жизни, каждый в отдельности предоставляют информацию для оценки исхода. Например, если нам предстоит выбрать между двумя методами лечения онкологического заболевания, один из которых обеспечивает 60% 5-летнюю выживаемость, а второй — 80%, и если показатель выживаемости является единственным критерием выбора, то последний метод будет предпочтительным, по сравнению с первым, поскольку он обеспечивает большую продолжительность жизни пациента. Тем не менее, такое сравнение совершенно не предполагает учёта качества жизни. Данные соображения особенно актуальны в ориентированной на пациента стратегии принятия решений в таких областях, как онкология, где поиск компромисса между про-

Вид измерения	Уровень оценки	Возможные пользователи информации, полученные на основе применения QALY				
		Пациент	Индивидуальный поставщик	Поставщик групп	Работодатель /страховщик	Государство
Мониторинг состояния здоровья	Отдельные пациенты					
	Группы пациентов					
	Популяции					
Сопоставление с нормативами и справочными данными	Отдельные пациенты					
	Группы пациентов					
	Популяции					
Определение изменений во времени	Отдельные пациенты					
	Группы пациентов					
	Популяции					

Рис. 1. Пользователи и области применения показателей состояния здоровья населения
Чёрная штриховка — качество жизни с поправкой на год (QALY) имеют минимальное отношение к процессу принятия решений разработчиками
Серая штриховка — QALY может (или не может) иметь отношение к принятию решений
Отсутствие штриховки — QALY, вероятно, имеет отношение к принятию решений.

должительностью и качеством жизни особенно важен. Именно по этой причине стандартизированные меры HRQoL являются такими многочисленными и широко используются длительное время.

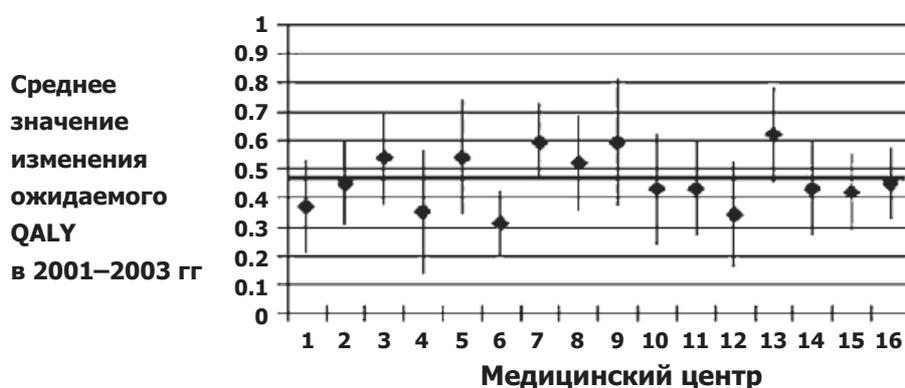
В пределах данной медицинской области не представляет особого труда убедить разработчиков медицинских технологий использовать QALY в качестве меры полезности для здоровья. Когда же влияние на выживаемость (к примеру) двух альтернативных хирургических манипуляций одинаково, например, при фиксации перелома бедра, любые различия HRQoL будут основой для сравнительной оценки QALY. Что же тогда будет иметь практическое значение для хирурга-ортопеда, которому предлагают использовать широкий потенциал основанной на QALY оценки для определения приоритетов вне области его специализации? Какую информацию получит он для себя или своих пациентов из данной наиболее общей оценки результатов вмешательств?

Несмотря на всё это, возможно, существуют дополнительные возможности для более широкого управления ролью QALY независимо от его первичного назначения для анализа полезности затрат. Например, недавно появились сообщения об опыте применения QALY в оценке улучшения качества жизни при сахарном диабете [22]. В рамках Контроля факторов риска сахарного диабета (программы по лечению диабета) [22] QALY использовали для сравнения показателей, полученных у пациентов с сахарным диабетом при проведении соответствующего мониторинга и лечения. Исторически сложилось так, что большинство связанных с диабетом показателей качества жизни были сосредоточены в нескольких измерениях, которые не были эмпириче-

ски связаны с улучшением исхода, или на отдельных степенях риска различных факторов, как правило, представленных в виде процентов от числа пациентов, взятых под «контроль». Подобный подход не только не позволяет лицам, принимающим решение обосновать выделении ресурсов для оказания помощи пациентам с сахарным диабетом, но также не дает возможности распределить ресурсы между несколькими направлениями в лечении диабета. Сводный показатель, основанный на стандартной оценке, потенциально значительно улучшит процесс принятия решений. Рис. 2 иллюстрирует возможности использования ожидаемого среднего изменения QALY в качестве индекса производительности в 16 медицинских центрах Kaiser Permanente.

Несмотря на то, что инвесторы и пациенты заинтересованы в раздельном учёте информации о качестве и продолжительности жизни для принятия решений, повышение актуальности QALY как комплексного показателя для этих групп пользователей зависит от возможностей его интерпретации. QALY был разработан для поэтапного обеспечения процессов принятия решений посредством CUA. Тем не менее, необходимы различные практические и теоретические фундаментальные установки для одновременного применения QALY как пациентами, так и врачами. Некоторые считают необходимым преодолеть этот разрыв и адаптировать критерий QALY для более широкого применения в клинической практике. Сюда относятся и методы, которые, как, например, HRQoL, находясь в руках пациента и врача в процессе принятия решений генерируют результат, выраженный мерой QALY, в соответствии с клинической ситуацией.

МИРОВЫЕ ПРИМЕРЫ ПРОЦЕССА РАЗРАБОТКИ РЕШЕНИЙ (САХАРНЫЙ ДИАБЕТ)



Среднее изменение QALY в медицинских центрах Kaiser Permanente

Заимствовано из Schmittziel et al, Predicted Quality-Adjusted Life Years as a Composite Measure of the Clinical Value of Diabetes Risk Factor Control, *Medical Care* Volume 45, Number 4, April 2007, 315–21.

Рис. 2. Процесс разработки решений по лечению сахарного диабета. QALY — продолжительность жизни, откорректированная по качеству

Очевидное отсутствие консенсуса между разработчиками инструментария HRQoL, методистами и экономистами препятствует восприятию QALY в качестве стандартной системы оценки. Критики или преуменьшают, или обходят вниманием QALY. Так, например, хотя общепризнанно, что разные методы определения утилитарности используют различные количественные критерии, нет чёткого соглашения, какой из них считать предпочтительным и, более того, существует ли такой критерий вообще. Специально данный вопрос не изучали, вследствие чего совершенно упускают из внимания, что существование стандартного метода возможно при условии развития концепции QALY. Аналогичное заключение можно сделать и в отношении выбора референтных значений для распределения ценностей. Предпочтение общественных ценностей над индивидуальными ценностями пациента (или учреждения здравоохранения) значения будет зависеть от системы воззрений пользователя QALY. Для лиц неэкономической специальности, обесценивание ценностей пациента трудно осознаваемо.

Выдвигаем на первый план QALY

Необходимо доказать важность независимого рассмотрения компонентов QALY и отделения, по существу, технических вопросов в оценке состояния здоровья. Методика создания базы данных пациентов для оценки качества жизни предвосхитила потребность Национальной службы здравоохранения Великобритании в 2009 году в пре- и послеоперационной оценке состояния здоровья пациентов, подвергшихся отдельным вмешательствам [23]. Использование таких данных в рамках национальной системы здравоохранения поможет облегчить интерпретацию информации, полученной с помощью QALY. Учёт HRQoL в повседневной клинической практике может оказаться в дальнейшем перспективным и для достижения других целей.

Рутинное применение HRQoL в оперативной и клинической практике само по себе будет способствовать акцентуации QALY. Включение QALY в процесс предоставления медицинской помощи, может изменить отношение пациентов к решениям, основанным на нем. Но должны быть разработаны своего рода рекомендации по использованию HRQoL, так как продолжающиеся дебаты о выборе системы оценки представляют существенную проблему, препятствуя распространению данного способа оценки и подрывая попытки продемонстрировать, чем данные HrQoL могут быть полезны.

Возвращаясь к лицам, принимающим решения, спросим у них, что нужно сделать и каков должен быть первый шаг для достижения реальных изме-

нений. Принятие разработчиков решений данного инструмента требует некоторого опыта обращения с ним. Это образовательная предпосылка подразумевает, в первую очередь, понимание того, как HrQoL можно использовать для принятия решений на уровне, ориентированном на пациента. Для акцентуации QALY, могут оказаться полезными такие подходы, как Транстеоретическая модель (ТТМ) изменений, разработанная Prokaska [24, 25]. ТТМ была успешно применена в исследованиях по принятию решений на уровне пациента [26]. Модель ТТМ предполагает, что исходно разработчики решений находятся на разных стадиях в их потенциальной способности и интереса к использованию QALY. Например, пациенты, вероятно, не знают о существовании QALY, а врачи могут быть осведомлены о QALY, но могут быть настроены к нему скептически или неоднозначно. Необходимы различные стратегии и просветительные усилия в зависимости от стадии по ТТМ, в которой находится на данный момент каждый из участников процесса.

Заключение

Итак, отмечен прогресс в использовании QALY в процессе принятия решений по вопросам распределения ресурсов в популяциях пациентов. В тоже время, в рамках практической деятельности понимание роли QALY развивается крайне медленно. Для более целостного восприятия QALY в качестве компонента процесса принятия решений за пределами рамок его традиционного понимания как меры экономической оценки, необходимо следующее. Во-первых, чтобы выйти за пределы требований рентабельности, что явно узаконит место QALY, требуются совместные усилия, чтобы строго определить сферы использования QALY в различных ситуациях при принятии решений. Для этого необходимо стремиться к улучшению как качества процесса принятия решений, так и его окончательного результата. Мы должны углубить наше понимание той степени, в которой лица, принимающие решения, будут использовать информацию, основанную на показателях QALY. Этот вопрос является наиболее актуальным для исследования. Недостаток данных по данному вопросу доказывает, что область эта мало изучена. Чтобы добиться прогресса в ней, мы должны сделать больше, чем доказать, что существует готовая аудитория для принятия такого рода информации. Единственный правильный путь состоит в исследованиях и идентификации ценностей реального мира клиницистов и пациентов. ISPOR должен играть ведущую роль в этом образовательном процессе, что позволит практиковать и руководить внедрением технологии принятия QALY.

Литература

1. American College of Physicians. Information on cost- effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med* 2008;148:956—61.
2. *Aspinall S.L., Good C.B., Glassman P.A., Valentino M.A.* The evolving use of cost-effectiveness analysis in formulary management within the Department of Veterans Affairs. *Med Care* 2005;43: 20—6.
3. *Parameswaran R., McNair A., Avery K.N., et al.* The role of health-related quality of life outcomes in clinical decision making in surgery for esophageal cancer: a systematic review. *Ann Surg Oncol* 2008;15:2372—9.
4. *Halyard M.Y., Ferrans C.E.* Quality-of-life assessment for routine oncology clinical practice. *J Support Oncol* 2008;6:221—33.
5. *Alishiri G.H., Bayat N., Fathi Ashtiani A., et al.* Logistic regression models for predicting physical and mental health-related quality of life in rheumatoid arthritis patients. *Mod Rheumatol* in press.
6. *Kolotkin R.L., Crosby R.D., Gress R.E., et al.* Health and health-related quality of life: differences between men and women who seek gastric bypass surgery. *Surg Obes Relat Dis* 2008;4: 651—8.
7. *Efficace F., Kemmler G., Vignetti M., et al.* Health-related quality of life assessment and reported outcomes in leukaemia randomised controlled trials—a systematic review to evaluate the added value in supporting clinical decision making. *Eur J Cancer* 2008;44: 1497—506.
8. Weinfurt KP. Outcomes research related to patient decision making in oncology. *Clin Ther* 2003;25:671—83.
9. *Kaplan R.M.* Shared medical decision-making: a new paradigm for behavioral medicine—1997 presidential address. *Ann Behav Med* 1999;21:3—11.
10. *Eiser C., Darlington A.S., Stride C.B., Grimer R.* Quality of life implications as a consequence of surgery: limb salvage, primary and secondary amputation. *Sarcoma* 2001;5:189—95.
11. *de Haes H, Koedoot N.* Patient centered decision making in palliative cancer treatment: a world of paradoxes. *Patient Educ Couns* 2003;50:43—9.
12. Solomon M.J., Pager C.K., Keshava A., et al. What do patients want? Patient preferences and surrogate decision making in the treatment of colorectal cancer. *Dis Colon Rectum* 2003;46: 1351—7.
13. *Protiere C., Viens P., Genre D., et al.* Patient participation in medical decision-making: a French study in adjuvant radio- chemotherapy for early breast cancer. *Ann Oncol* 2000;11:39—45.
14. *Schackman B.R., Teixeira P.A., Weitzman G., et al.* Quality-of-life tradeoffs for hepatitis C treatment: do patients and providers agree? *Med Decis Making* 2008;28:233—42.
15. *Goodlin S.J., Quill T.E., Arnold R.M.* Communication and decision-making about prognosis in heart failure care. *J Card Fail* 2008;14:106—13.
16. *Pruchno R.A., Lemay E.P. Jr, Field L., Levinsky N.G.* Predictors of patient treatment preferences and spouse substituted judgments: the case of dialysis continuation. *Med Decis Making* 2006;26: 112—21.
17. *Grosset K.A., Grosset D.G.* Patient-perceived involvement and satisfaction in Parkinson's disease: effect on therapy decisions and quality of life. *Mov Disord* 2005;20:616—9.
18. *Hankins J., Hinds P., Day S., et al.* Therapy preference and decision-making among patients with severe sickle cell anemia and their families. *Pediatr Blood Cancer* 2007;48:705—10.
19. *Sullivan M.D., Levy W.C., Russo J.E., et al.* Summary health status measures in advanced heart failure: relationship to clinical variables and outcome. *J Card Fail* 2007;13:560—8.
20. *Glasziou P., Alexander J., Beller E., Clarke P.* Which health-related quality of life score? A comparison of alternative utility measures in patients with Type 2 diabetes in the ADVANCE trial. *Health Qual Life Outcomes* 2007;5:21.
21. *Watkins J.B., Minshall M.E., Sullivan S.D.* Application of economic analyses in U.S. managed care formulary decisions: a private payer's experience. *J Manag Care Pharm* 2006;12:726—35.
22. *Schmittiel J., Vijan S., Fireman B., et al.* Predicted quality-adjusted life years as a composite measure of the clinical value of diabetes risk factor control. *Med Care* 2007;45:315—21.
23. Department of Health. Guidance on the routine collection of Patient Reported Outcome Measures (PROMs). Available from: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_091451 [Accessed December 12, 2008].
24. *Prochaska J.M., Prochaska J.O., Levesque D.A.* A transtheoretical approach to changing organizations. *Adm Policy Ment Health* 2001;28:247—61.
25. *Prochaska J.O., Velicer W.F.* The transtheoretical model of health behavior change. *Am J Health Promot* 1997;12:38—48.
26. *Levesque D.A., Cummins C.O., Prochaska J.M., Prochaska J.O.* Stage of change for making an informed decision about medicare health plans. *Health Serv Res* 2006;41:1372—91.

Слово к Консенсусу по QALY

Michael Drummond, DPhil¹, Diana Brixner, BPharm, PhD², Marthe Gold, MD, MPH³, Paul Kind⁴, Alistair McGuire, PhD⁵, Erik Nord, PhD⁶ (Consensus Development Group)

¹ — University of York, York, UK

² — University of Utah, Salt Lake City, Utah, USA

³ — City University of New York Medical School, New York, NY, USA

⁴ — University of York, York, UK

⁵ — London School of Economics, London, UK

⁶ — Norwegian Institute of Public Health, Oslo, Norway

Переводчик:

Вилюм Ирина Александровна — ассистенту, Первый Санкт-Петербургский государственный медицинский университет им. акад. И. П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Белоусов Дмитрий Юрьевич — генеральный директор ООО «Центра фармакоэкономических исследований», г. Москва

Ключевые слова: распределение ресурсов, качество жизни, оценка состояния здоровья

Toward a Consensus on the QALY

Keywords: resource allocation, quality of life, health state valuation

Введение

Материалы предыдущих статей Специального выпуска «Экономической оценки в здравоохранении» — «Выдвигаем QALY на первый план: строим прагматический путь» (статьи 2—6) были представлены на Семинаре по развитию ISPOR с аналогичной темой, который состоялся в Филадельфии 6-8 ноября 2007 года. После обсуждения статей, была сформирована рабочая группа по развитию консенсуса, в состав которой вошли представители авторского коллектива каждой статьи. Это было необходимо для выявления точек соприкосновения по ключевым вопросам.

Учитывая разнообразие мнений, высказанных на этом семинаре, как следует из представленных статей, стало очевидно, что достичь согласия в специфических вопросах, касающихся, к примеру, способов измерения QALY, или вспомогательных инструментов для его оценки, будет крайне тяжело. Скорее, Группа сочла, что самое верное решение заключается в достижении соглашения по нескольким принципиальным глобальным вопросам и обозначении противоречий в виде набора вопросов, подлежащих дальнейшему исследованию. В целом, восемь основных постановлений Семинара были вполне согласованы, что определило

множество дискуссий среди участников. Эти постановления изложены ниже, наряду с основными вопросами, вызвавшими дискуссию.

Постановления Консенсуса

QALY представляет собой один из обоснованных подходов к оценке здоровья и разработке решений в здравоохранении

Это утверждение содержит два важных момента. Во-первых, QALY, в нынешнем звучании, концентрируется на здоровье, а не на благополучии в целом. Это является его сильной стороной, учитывая, что основной целью большинства систем здравоохранения является улучшение именно здоровья населения.

Во-вторых, данным постановлением признается, что есть и другие подходы к разработке решений в здравоохранении, которые должны быть рассмотрены вместе с QALY. Они не были рассмотрены в деталях, но, по данным литературы, можно утверждать, что равенство и социальная справедливость также являются определяющими принципами при распределении ресурсов здравоохранения. В ходе обсуждения было отмечено, что для уточнения сути и роли данных принципов требуются дальнейшие исследования (см. постановление 5).

В системе медицинского обслуживания QALY может использоваться на различных уровнях

Во время конференции выяснилось, что, по мнению участников, QALY потенциально способствует принятию различных типов решений.

Традиционно, QALY применяли для обеспечения процесса распределения ресурсов среди различных групп населения. С другой стороны, некоторые участники предложили использовать QALY для определения способов лечения отдельных пациентов или их групп. В контексте частной системы здравоохранения, QALY мог бы быть использован для обоснования решений о страховом покрытии или уровнях доплат.

Важно обозначить данные возможности использования QALY, так как единственным источником разногласий остаётся контингент, по которому учитываются значения для вычисления QALY. Так, например, в связи с тем, что QALY часто используют для разработки популяционных решений, обычно в качестве целевого контингента для выявления предпочтений по здоровью выбирают представителей общей популяции. С другой стороны, для принятия клинических решений, с определением оптимального способа лечения для различных состояний, как правило, используют данные, полученные при опросе непосредственно пациентов.

В ходе обсуждения было отмечено, что, несмотря на определенный опыт применения QALY в разработке решений по распределению ресурсов (к примеру, работа, проведенная в Национальном Институте здравоохранения и качества медицинской помощи (NICE) в Великобритании), остаются мало исследованными другие области для принятия решений с помощью QALY. Таким образом, роль QALY, как универсального инструмента для принятия решений, должна стать приоритетным направлением для исследований, возможно, методом ситуационного анализа и в различных условиях. Участники конференции отметили, что в странах, где правительство не является основным источником финансирования здравоохранения, решения о распределении ресурсов имеют менее широкую базу, что рождает необходимость определить, как лиц, принимающих решения в таких условиях, так и способы для наилучшего применения QALY.

Также в ходе дискуссии было отмечено, что, несмотря на заявления о необходимости межотраслевой правительственной инициативы, относительно мало известно о взаимосвязи между состоянием здоровья и благополучия. Более того, мало известно о побочных эффектах различных медицинских вмешательств, а также о последствиях практических мероприятий в области здравоохранения, программ служб социальной защиты, образования, или защиты окружающей среды.

Если цель состоит в оптимизации использования всех ресурсов в рамках общественных программ, понадобится разработка расширенной модели QALY, как меры благосостояния. Тем не менее, это остаётся вопросом для дальнейшего исследования и не должно заставлять нас пренебрегать QALY как важным компонентом текущего процесса принятия решений в здравоохранении.

Необходимо учитывать как предпочтения, полученные с опытом, так и *ex ante*

Одним из основных дискуссионных вопросов во время семинара заключался в том, могут ли лица, не находившиеся в определённых состояниях здоровья, адекватно оценивать их (это обсуждение перекликается с предметом дискуссии, состоявшейся на 11-й Ежегодной международной встрече ISPOR в мае 2006 г.).

Этот вопрос непосредственно связан с другим, который обсуждался в рамках 2-го постановления Консенсуса, о том, что представители широкой общественности едва ли могли испытать все состояния здоровья, служащие предметом оценки, однако они могут иметь друзей или родственников, перенесших их. Таким образом, для некоторых исследователей допустимо принимать во внимание данные по общей популяции в том случае, если они заняты глобальным распределением ресурсов, в то же время следует согласиться с приоритетом данных, полученных от лиц, испытавших те или иные состояния здоровья, ввиду лучшего их понимания проблемы.

Рабочая группа отметила несколько нерешенных вопросов относительно измерения экспериментальной полезности. Так, например, в исследованиях по оценке полезности среди пациентов и инвалидов, многие авторы сообщают о нежелании респондентов жертвовать хоть сколько-нибудь продолжительностью жизни ради клинического улучшения, даже при наличии у них значительных проблем со здоровьем. Поэтому, несмотря на важность данной информации для разработки решений по организации медико-санитарной помощи, до сих пор не ясно, как её можно выразить в единицах принятой шкалы от 0 до 1 [1].

Тем не менее, Консенсус постановил, что обоснован учет предпочтений, как высказанных *ex ante*, лицами, никогда не испытывавшими данное состояние, так и предпочтений, основывающихся на опыте нахождения в определённом состоянии, в зависимости от задач разработчиков и принятых нормативов. В некоторых случаях, возможен сочетанный учёт предпочтений, при котором члены общей популяции могут быть проинформированы о точке зрения лиц, непосредственно перенесших состояние здоровья, служащее предметом оценки.

Рабочая группа не акцентировала внимание на технологических вопросах оценки предпочтений, однако отметила необходимость дальнейших исследований

индивидуальных предпочтений пациентов, их влияния на общественные взгляды и возможности применения для информирования широкой общественности.

Необходимо ставить вопросы распределения

Отношение цена/QALY отражает экономическую эффективность вмешательства. Следовательно, данное отношение есть скорее мера полезности, нежели справедливости. Проблемы учета понятий справедливости (другими словами, относительность предпочтений по продолжительности и качеству жизни, высказанных индивидами с разным состоянием здоровья и различным потенциалом к выздоровлению), которые не принимаются во внимания при расчете QALY, вызвали бурные обсуждения на семинаре. В процессе распределения ресурсов в группах больных эти вопросы звучат в контексте эффективности для здоровья. Учет данных соображений может существенно сместить общественные приоритеты в распределении ресурсов в сторону от отношения стоимость/ QALY. Данные проблемы имеют международное значение во всех сферах исследования и/или применения анализа экономической эффективности.

Рабочая группа обозначила три принципиально возможных пути решения проблемы распределения при оценке QALY. Во-первых, сам метод определения QALY должен включать поправку на распределение. К примеру, метод персональных уступок, в отличие от других методов, тем и отличается от остальных методов, что учитывает индивидуальные различия в предпочтениях лиц, имеющих состояния разной степени тяжести. Для включения в методику оценки QALY данных поправок методом персональных уступок и другими методами уже имеются как теоретические предпосылки, так и практический опыт.

Во-вторых, QALY, вычисляемый с помощью оценочных шкал, таких как визуальная аналоговая шкала (VAS), временных уступок (ТТО), или стандартного риска (SG) может быть взвешен в соответствии с распределением социальной справедливости, а не являться постоянной величиной, как это имеет место в настоящее время. Процесс «взвешивания» может основываться на исследовании предпочтений представителей популяции, или быть разработан с помощью группы присяжных. Данная работа будет включать поиск нормативных факторов (например, степень тяжести состояния, способность к выздоровлению, справедливость очередности), которые будут иметь приоритет при распределении весов.

В-третьих, QALY может быть сохранен в качестве меры производительности здравоохранения, тогда он включит распределительные вопросы посредством совещательного процесса разработки решений, который будет сопровождать процесс распределения ресурсов. Например, комитеты по разработке решений о выделении ресурсов на здравоохранение,

руководствуясь коэффициентами справедливого распределения, могли бы повысить экономическую эффективность некоторых вмешательств.

Хотя рабочая группа не пришла к единому мнению по предпочтительным подходам, были определены некоторые ключевые соображения для выбора. С одной стороны, существует потребность в четкости понимания QALY. Усложняя QALY путем включения системы весов, мы затрудняем его интерпретацию. В условиях, когда основным препятствием для успешного использования QALY является отсутствие понимания его со стороны разработчиков решений, усложнение его структуры лишь ухудшит ситуацию.

С другой стороны, количественная оценка с учетом распределительных коэффициентов может оказаться полезной для расчета общей стоимости вмешательств в здравоохранении. Если эти вопросы для разработчиков являются второстепенными наряду с расчетом стоимости, они не могут быть адекватно рассмотрены.

Существует потребность в научных исследованиях по сравнению различных подходов для включения распределительных аспектов в процесс принятия решений по распределению ресурсов. К примеру, как учет предпочтений *ex ante* может служить для оценки QALY с помощью метода персональных уступок? Какую роль должны играть дистрибутивные вопросы: быть принятыми во внимание при фактической количественной оценке общего QALY (расчет равновесных значений), или служить дополнением в процессе принятия решений? Насколько реально разработать соответствующие равновесные значения?

Всеобщее признание на семинаре получил тот факт, что дистрибутивные вопросы занимают центральное положение при распределении ресурсов, но при отсутствии эмпирических данных о наиболее удобном для разработчиков подходе, согласие в этом вопросе так и не было получено. В настоящее время статус-кво сохраняет дистрибутивные вопросы за пределами понятия экономической эффективности, и многие участники семинара высказались в поддержку приоритета традиционного QALY. Тем не менее, существует настоятельная необходимость привлечения внимания разработчиков к более эффективному использованию дистрибутивных аспектов в экономической оценке.

Различные методы оценки в здравоохранении дают различные результаты, что требует более глубокого изучения

Множество методов выявления предпочтений по состояниям здоровья принципиально делятся на две группы: 1) описательные системы; и 2) оценочные методики (например, ТТО, SG, VAS, РТО). Процесс характеристики состояния здоровья является важной предпосылкой для установления твердых оце-

нок. Хотя различные описательные системы имеют много общего, между ними существуют и важные различия (к примеру, Индекс полезности здоровья (HUI) не учитывает роль социального положения, а модель EQ-5D — уровень напряжения).

Взаимосвязь между различными подходами и обоснованностью их применения подлежит дальнейшему изучению. Более того, необходимо понять, в какой степени согласуются описательные системы и соответствующие системы количественной оценки, насколько они стабильны во времени и насколько специфично для разных направлений в медицине. Также, должны быть изучены эквиваленты перехода от одной описательной системы к другой (см. далее).

Основным назначением QALY служит оценка стоимости вмешательств с позиций полученной прибыли (или предотвращенных убытков) в области здравоохранения. В традиционной методике расчёта QALY эта оценка является косвенной: на основе предпочтений здоровых людей *ex ante* производится оценка состояний здоровья и соответствующей разницы между состояниями до и после вмешательства, так называемого коэффициента усиления («метод вычитания»). Основным недостатком данного подхода является учет оценки лиц, не имеющих заболевания, такая оценка стоимости лечения может оказаться ниже, чем пропорциональная величина эффекта от лечения [1]. Этот вопрос также требует изучения. Аналогичным образом, можно подвергнуть сомнению предположение (не напрямую связанное с расчетом QALY) о пропорциональности числа лет жизни и их стоимости. В этой области, как и во многих других, действует закон убывающей предельной полезности товаров (в данном случае: лет).

Прирост здоровья необходимо уравнивать по времени его достижения

Рабочая группа Консенсуса признает, что традиционный подход, предполагающий аккумуляцию эффектов для здоровья во времени довольно упрощён; а именно, значение, соответствующее определённому состоянию здоровья, умножается на время, проведенное в каждом состоянии. Необходимы дополнительные исследования линейной зависимости предпочтений от времени, а также разработка специальных шкал и профилей. Рабочая группа обращалась к опыту дискуссии об «эквивалентах здоровых лет» (HUEs) от 1980-х г., когда было признано, что HUEs представляют собой принципиально наилучший подход, но трудно применимый, в частности, в контексте изменения состояний здоровья в ходе оказания медико-санитарной помощи. Таким образом, вопросом не только практическим, но и теоретическим остаётся разработка методов уравнивания выгод для здоровья по времени, затраченному на их получение.

Необходим «Контрольный метод» для валидации QALY

Предпосылкой для возникновения данного постановления послужило то, что рабочая группа отметила, что QALY еще не полностью не раскрыл свой потенциал, в особенности, в аспекте его применения для разработки решений. Однако, учитывая количество имеющихся разногласий относительно предпочтительного подхода в оценке, группа полагает, что необходимо сравнение с неким «эталонным методом», впервые разработанным в США Группой специалистов по экономической эффективности в здравоохранении и медицине [2], что сопоставить значение обоих методов для применения в контексте различных заболеваний и вмешательств. Рассматривая референсный метод, необходимо использовать стандартизированный подход также и для оценки QALY, чтобы модель была пригодна для всех вариантов экономической оценки, без исключения. Это будет способствовать сопоставимости экономических оценок, не препятствуя дальнейшему совершенствованию методологии QALY.

Участники Консенсуса убеждены в том, что данный шаг необходимо предпринять для завершения дебатов. В противном случае, подобные семинары будут проходить еще 20 лет. Тем не менее, члены рабочей группы не рекомендовали ни один метод в качестве референсного, отметив, что 1) каждый может выбрать показатель, или комплекс показателей, и принять их; 2) каждый может провести процедуру «перекрестного сопоставления» (или трансляции) между наиболее распространенными показателями, таким образом, что, в принципе, любой метод будет применим [3]; или 3) каждый вправе создать группу, включающую как методистов, так и разработчиков решений, которая рассмотрит варианты показателей и выделит наилучший.

Рабочая группа признала, что поиск эталонного метода таит в себе потенциальные методологические и практические риски. К примеру, все существующие показатели (такие как EQ-5D, HUI) базируются на конкретных атрибутах состояния здоровья / нездоровья. Таким образом, избрав один из показателей, мы неминуемо смещаем фокус глобального распределения ресурсов здравоохранения в пользу тех мероприятий, которые максимально влияют на его атрибуты. Это смещение может быть сведено к минимуму путём разработки базового подхода, который был бы основан на нескольких показателях, но это более дорогостоящий подход (так как требует набора первичных данных, в ходе клинических испытаний). Однако, несмотря на признание описанных рисков, участники консенсуса не нашли альтернативного выхода для дальнейшего развития.

Как и следовало ожидать, это постановление вызвало немало дискуссий. Несколько участников семинара не согласились, что выбор эталонного ме-

тогда необходим для дальнейшей работы. Другие считали, что, при определении эталонного подхода, не обязательно ограничивать себя набором существующих показателей, каждый из которых имеет свои недостатки. Некоторые считали, что следует дождаться результатов проводимых перекрестных исследований [3] прежде, чем сделать следующий шаг.

Заключение

Постановления Консенсус, приведенные выше, представляют собой соглашение по ряду «горячих» вопросов. Большинство из них были приняты практически единогласно, хотя в ряде вопросов требовались уточнения. Наибольшие разногласия вызвало последнее постановление, когда несколько участников не смогли одобрить необходимость разработки эталонного метода для оценки QALY на данный момент времени. Подавляющее большинство, однако, высказалось в пользу данного шага, который представляется возможным путем для совершенствования применения QALY при анализе экономической эффективности.

В ходе обсуждения основных вопросов и последующих дискуссий среди участников семинара был составлен список вопросов, которые могут служить предметом для дальнейших исследований. Они приведены в табл. 1.

Обсуждение: ищем пути вперед

Сформулировав ряд постановлений и сгенерировав список вопросов для дальнейших исследований, дискуссия коснулась некоторых практических вопросов, подлежащих решению для совершенство-

вания QALY. Многие из них вращались вокруг настоящей и будущей разработки решений для здравоохранения по обе стороны Атлантики. Избранные вопросы были связаны с проблемами «операционного характера», в частности разработкой эталонного метода для оценки QALY.

Применение QALY в современных условиях

Было высказано суждение о том, что следует отдавать предпочтение наблюдению за работой таких организаций, как, в частности, NICE в Великобритании, которые в настоящее время используют QALY [4]. Например, есть ли доказательства того, что это позволяет принимать «лучшие» решения? Считают ли сами сотрудники NICE себя состоятельными для решения данной задачи? Что ещё важно учитывать в процессе принятия решений наряду с показателями прироста относительной рентабельности (ICER)? Каков опыт участия граждан в процессе принятия решений в составе Совета Граждан NICE?

Проблемы:

использование QALY в избранных ситуациях

Как уже было отмечено, большинство пользователей QALY являются членами централизованной, финансируемой из единого фонда, системы здравоохранения, такой как Британская национальная Служба здравоохранения. Несколько участников Консенсуса выразили сомнение о значении QALY в децентрализованных системах здравоохранения, таких как в Соединенных Штатах, где улучшение здоровья населения не является главным приоритетом (а зависит от индивидуальных возможностей).

Таблица 1

Области дальнейших исследований по применению QALY

1.	Применение экономических задач к разработке решений для здравоохранения
2.	Наблюдения применения QALY на разных уровнях системы здравоохранения
3.	Наблюдения опыта применения QALY в децентрализованных, финансируемых из частных фондов системах оказания медицинской помощи
4.	Вклад составляющей здоровья в широкое понятие благосостояния
5.	Роль расширенной модели QALY, включающей не только показатели здоровья
6.	Соотношение оценок, полученных от лиц, находящихся или находившихся в том или ином состоянии здоровья, и оценок, собранных в общей популяции, по этим же состояниям
7.	Методы информирования членов общей популяции об опыте пациентов, перенесших то или иное состояние
8.	Сопоставление основных методов оценки состояния здоровья с учетом дистрибутивных аспектов
9.	Анализ поправочных дистрибутивных коэффициентов
10.	Качественные исследования общественного мнения в отношении дистрибутивных вопросов
11.	Оценка полезности для здоровья и состояний здоровья
12.	Исследование линейной зависимости предпочтений от времени и поиск способов получения шкал соответствия и профилей
13.	Разработка эталонного метода или серии эталонных методов для сопоставления с QALY

Однако, другие участники отметили, что основные плательщики (к примеру, WellPoint), и крупные управляемые медицинские организации (такие как Kaiser-Permanente) охватывают «сообщество» отдельных лиц, которые по численности могут сравниться с численностью населения многих европейских странах. Некоторые из этих организаций уже определили степень воздействия терапии на качество жизни в процессе принятия решений [5, 6]. Было также отмечено, что, хотя в прошлом необходимость экономического анализа и встречала сопротивление, настроение в Соединенных Штатах претерпевало изменения по мере снижения доступности медицинской помощи как в государственном, так и частном секторе здравоохранения. Тогда членами широкой общественности [7] и разработчиками решений [8, 9] была признана необходимость учета ограничений системы здравоохранения. Данные ограничения явились предметом нынешних дискуссий о новой американской федеральной инициативе по проведению сравнительных исследований эффективности различных мероприятий в здравоохранении, дополненных анализом эффективности затрат [10].

Участники семинара выделили следующие направления для дальнейшей деятельности: 1) выяснить требования разработчиков решений по использованию QALY и экономического анализа; 2) информировать их в отношении преимуществ и ограничений метода; и 3) провести наблюдения процесса разработки решений по различным направлениям, для подбора оптимальных методов. Матрица для разработки решений в здравоохранении, приведённая в предыдущей статье Специального выпуска, будет служить полезной отправной точкой для выявления ситуаций, в которых требуются эти исследования.

Разработка эталонного метода оценки QALY

Так как большинство участников семинара высказались в пользу данной разработки, предметом дискуссии явился путь осуществления данной задачи. Во-первых,

Кто должен руководить данной разработкой? Должна ли она осуществляться международными усилиями, во главе с такой организацией, как ISPOR, или

должна опираться на серию региональных исследований, учитывая различные потребности разработчиков решений и различные традиции (как, например, в Соединенных Штатах и Европе) в использовании тех или иных инструментов? Были предложены следующие ответственные органы в США: Агентство по охране здоровья и качеству медицинской помощи, Медицинский Институт и Национальный Институт Здоровья. Данный вопрос не был решён окончательно.

Во-вторых, то, кто должен войти в состав группы по разработке эталонного подхода? Общее мнение состояло в том, что состав данной группы должен быть как можно шире и включать экономистов, эпидемиологов, разработчиков решений по общественному здоровью, служащих и управляющих в здравоохранении, членов правительства.

Наконец, необходимо определить временные рамки? Несколько участников предположили, что, если необходимость получения дополнительных данных будет служить предлогом для отказа от разработки эталонного метода, он никогда не будет создан. Консенсус состоял в том, что действия должны предприниматься безотлагательно, при условии пересмотра любых решений через несколько лет.

Заключение

Семинар был успешным по двум причинам. Во-первых, широкому обсуждению были подвергнуты перспективы использования QALY.

Во-вторых, был достигнут консенсус по ряду принципиальных вопросов и разработана программа дальнейших исследований.

Однако, обозначить единственный путь для развития представлялось мало вероятным, что, учитывая разнообразие взглядов, не удивительно. Тем не менее, по мнению большинства, QALY должен более широкое признание со стороны разработчиков решений, необходимо предпринимать усилия по выдвиганию данного вопроса на первый план. По мнению большинства, основные усилия должны быть направлены на развитие эталонного метода для оценки QALY. Если это в конечном итоге произойдет, настоящий семинар будет поворотной точкой в развитии QALY.

Литература

1. Nord E., Daniels N., Kamlet M. QALYs: some challenges. *Value Health* 2008;12(Suppl. 1):S10—5.
2. Gold M.R., Siegel J.E., Russell L.B., Weinstein M.C., eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press, 1996.
3. Fryback D.G., Paltal M., Cherepov D., et al. Cross-walks among five self-reported summary health utility indexes: progress and prospects. *Med Decis Making* 2008;28:E24.
4. Rawlins M.D., Culyer A.J. National institute for clinical excellence and its value judgments. *BMJ* 2004;329:224—7.
5. Aspinall S.L., Good C.B., Glassman P.A., Valentino M.A. The evolving use of cost-effectiveness analysis in formulary management within the Department of Veteran Affairs. *Med Care* 2005;43(Suppl. 7):S20—6.
6. Wu W.K., Sause R.B., Zacker C. Use of health-related quality of life information in managed care formulary decision-making. *Res Social Adm Pharm* 2005;1:579—98.
7. Gold M.R., Sofaer S., Siegelberg T. Medicare and CEA; time to ask the taxpayers. *Health Aff (Millwood)* -2007;5:1399—406.
8. Sofaer S., Bryan S., Gold M.R. Decision-makers' attitudes toward cost-effectiveness analysis. Abstract. *Academy Health*, Orlando, July 5, 2007.
9. Neumann P.J., Palmer J.A., Daniels N., et al. Panel on integrating cost-effectiveness considerations into health policy decisions. *Am J Manag Care* 2008;14:185—8.
10. American College of Physicians. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med* 2008;149:956—61.

Анализ влияния на бюджет: принципы надлежащей практики. Доклад рабочей группы ISPOR по надлежащей практике анализа влияния на бюджет II, 2012 год

Sean D. Sullivan, PhD¹, Josephine A. Mauskopf, PhD², Federico Augustovski, MD, MSc, PhD³,
J. Jaime Caro, MDCM, FRCPC, FACP⁴, Karen M. Lee, MA⁵, Mark Minchin, MBA⁶,
Ewa Orlewska, MD, PhD^{7,8}, Pete Penna, PharmD⁹,
Jose-Manuel Rodriguez Barrios, RPh, MPH, MSc¹⁰, Wen-Yi Shau, PhD, MD¹¹

¹ — *Pharmaceutical Outcomes Research and Policy Program, University of Washington, Seattle, WA, USA*

² — *RTI Health Solutions, Research Triangle Park, NC, USA*

³ — *Health Economic Evaluation and Technology Assessment, Institute for Clinical Effectiveness and Health Policy, Buenos Aires, Argentina*

⁴ — *Evidera, Lexington, MA, USA*

⁵ — *Health Economics, Canadian Agency for Drugs and Technologies in Health, Ottawa, ON, Canada*

⁶ — *PASLU, National Institute for Health and Care Excellence, Manchester, UK*

⁷ — *Centre for Pharmacoeconomics, Warsaw, Poland*

⁸ — *The Jan Kochanowski University, Kielce, Poland*

⁹ — *Formulary Resources, LLC, Mercer Island, WA, USA*

¹⁰ — *Stryker, Madrid, Spain*

¹¹ — *Center for Drug Evaluation, Taipei, Taiwan*

Переводчики:

Галанкин Тимофей Леонидович — к.м.н., ассистент, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Мосикян А.А. — студентка 4 курса, лечебный факультет, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Рецензент:

Колбин Алексей Сергеевич — д.м.н., профессор, Первый Санкт-Петербургский государственный медицинский университет им. акад. И.П. Павлова, Санкт-Петербург, Российская Федерация

Резюме. Предпосылки: Анализ влияния на бюджет (Budget impact analysis, далее — BIA) является неотъемлемой частью комплексной экономической оценки интервенций в здравоохранении и всё чаще требуется финансовым органам, компенсирующим затраты на лечение пациентов (далее — регулятору), как для составления списков медицинских технологий, подлежащих компенсации, так и для утверждения документов на возмещение затрат. *Цели.* Целью этого доклада является представление обновленного методического руководства как для специалистов, выполняющих такого рода анализ, так и для пользователей отчетов выполненного анализа. Данное обновление стало необходимым отчасти из-за постоянного развития методов BIA, а отчасти из-за растущего к ним интереса, особенно

на развивающихся рынках при решении вопросов ценовой политики в здравоохранении, доступности медицинской помощи и ее влиянии на здоровье населения. *Методы.* Рабочая группа была одобрена Советом ISPOR по принятию решений в области медико-санитарных исследований и утверждена его Правлением. Членами группы являлись опытные разработчики или пользователи ВИА, работники научной сферы и промышленности, а также консультанты при правительствах, прибывшие из ряда стран Северной и Южной Америки, Океании, Азии и Европы. Рабочая группа ISPOR ходатайствовала о рецензировании доклада как группой внешних (основных) рецензентов, так и членами других подразделений ISPOR. *Результаты.* Рабочая группа высказалась в пользу того, что дизайн ВИА новых вмешательств (интервенций) в сфере здравоохранения должен учитывать локальные особенности системы здравоохранения, возможную ограниченность доступа к медицинским технологиям, частоту использования и эффективность текущих (стандартных) и новых медицинских интервенций, делать поправку на ожидаемое внедрение новых интервенций. Ключевые элементы анализа включают в себя оценку размера исследуемой (целевой) популяции, используемого в настоящий момент набора лечебных вмешательств (стандартной терапии), а также изменения после возможного внедрения в практику нового вмешательства. Анализ должен учитывать любые ожидаемые изменения в расходах, связанных с состоянием больного. По возможности, расчеты ВИА должны быть выполнены в виде простого калькулятора на базе Excel, чтобы регулятор не испытывал затруднений в их использовании. В тех случаях, когда изменения в размерах целевой популяции, сложная структура степеней тяжести состояния или методов лечения не могут быть достоверно отражены с помощью возможностей Excel-калькулятора, для оценки влияния на бюджет новой интервенции можно использовать когортные или субъектно-ориентированные модели, надлежащим образом учитывающие входящих и выходящих субъектов, а также тяжесть их состояния и объем требуемых вмешательств, на время горизонта моделирования. ВИА должен использовать данные, отражающие стоимости, специфичные для каждой конкретной популяции, интересующей регулятора. Анализ чувствительности для альтернативных сценариев также должен исходить из интересов и перспектив регулятора. Валидизация модели должна включать как минимум оценку внешней валидности, совместно с регулятором, и проверку качества выполненных расчетов. Источниками информации для ВИА служат опубликованные рандомизированные клинические испытания (РКИ), сравнительные исследования эффективности и безопасности стандартных и новых медицинских вмешательств. Остальные необходимые показатели можно получить непосредственно из статистик целевой популяции. Источниками информации могут являться опубликованные данные, общеизвестные локальные или национальные статистики, а также, в особых случаях, мнение экспертов. Отчет ВИА должен предоставлять подробную информацию о значениях входных параметров и о самих расчетах так, чтобы этот анализ мог быть воспроизведен другим исследователем. Результаты ВИА должны быть представлены в формате, удобном для восприятия регулятором. В готовом калькуляторе всегда должны представляться варианты различных ценовых категорий, которые можно произвольно включать или не включать в анализ. *Заключение.* Мы разработали рекомендации для структуры ВИА, а также руководство по сбору и использованию данных и предлагаем общий формат выполнения отчета, который будет способствовать стандартизации и прозрачности анализа. Соблюдение этих принципов надлежащей научно-исследовательской практики не должно вытеснять существующие методы проведения ВИА, но должно поддерживать и усилить локальные руководства и рекомендации, а также послужить отправным пунктом для регулятора, в случае возникновения у него желания способствовать распространению методических руководств.

Ключевые слова: анализ влияния на бюджет, калькулятор расчёта стоимости, экономическая оценка, методология, моделирование

Budget impact analysis: principles of good practice. Report of the ISPOR Working Group on Good Practices for Budget Impact Analysis II, 2012

Abstract. Background. Budget impact analyses (BIAs) are an essential part of a comprehensive economic assessment of a health care intervention and are increasingly required by reimbursement authorities as part of a listing or reimbursement submission. **Objectives:** The objective of this report was to present updated guidance on methods for those undertaking such analyses or for those reviewing the results of such analyses. This update was needed, in part, because of developments in BIA methods as well as a growing interest, particularly in emerging markets, in matters related to affordability and population health impacts of health care interventions. **Methods.** The Task Force was approved by the International Society for Pharmacoeconomics and Outcomes Research Health Sciences Policy Council and appointed by its Board of Directors. Members were experienced developers or users of BIAs; worked in academia and industry and as advisors to governments; and came from several countries in North America and South America, Oceania, Asia, and Europe. The Task Force solicited comments on the drafts from a core group of external reviewers and, more broadly, from the membership of the International Society for Pharmacoeconomics and Outcomes Research. **Results.** The Task Force recommends that the design of a BIA for a new health care intervention should take into account relevant features of the health care system, possible access restrictions, the anticipated uptake of the new intervention, and the use and effects of the current and new interventions. The key elements of a BIA include estimating the size of the eligible population, the current mix of treatments and the expected mix after the introduction of the new intervention, the cost of the treatment mixes, and any changes expected in condition-related costs. Where possible, the BIA calculations should be performed by using a simple

cost calculator approach because of its ease of use for budget holders. In instances, however, in which the changes in eligible population size, disease severity mix, or treatment patterns cannot be credibly captured by using the cost calculator approach, a cohort or patient-level condition-specific model may be used to estimate the budget impact of the new intervention, accounting appropriately for those entering and leaving the eligible population over time. In either case, the BIA should use data that reflect values specific to a particular decision maker's population. Sensitivity analysis should be of alternative scenarios chosen from the perspective of the decision maker. The validation of the model should include at least face validity with decision makers and verification of the calculations. Data sources for the BIA should include published clinical trial estimates and comparator studies for the efficacy and safety of the current and new interventions as well as the decision maker's own population for the other parameter estimates, where possible. Other data sources include the use of published data, well-recognized local or national statistical information, and, in special circumstances, expert opinion. Reporting of the BIA should provide detailed information about the input parameter values and calculations at a level of detail that would allow another modeler to replicate the analysis. The outcomes of the BIA should be presented in the format of interest to health care decision makers. In a computer program, options should be provided for different categories of costs to be included or excluded from the analysis. *Conclusions.* We recommend a framework for the BIA, provide guidance on the acquisition and use of data, and offer a common reporting format that will promote standardization and transparency. Adherence to these good research practice principles would not necessarily supersede jurisdiction-specific BIA guidelines but may support and enhance local recommendations or serve as a starting point for payers wishing to promulgate methodology guidelines.

Keywords: budget impact analysis, cost calculator, economic evaluation, methodology, modeling

Предыстория создания рабочей группы ISPOR

Данный доклад рабочей группы ISPOR представляет собой новую редакцию доклада предыдущей рабочей группы по BIA, опубликованного в 2007 году [1]. Дополнения стали необходимы, отчасти, из-за всё возрастающего признания BIA как самостоятельной экономической экспертизы, а также из-за растущего к нему интереса, особенно на развивающихся рынках, связанного с ценовой доступностью медицинской помощи и её влиянием на здоровье населения. Приведем перечень вопросов, рассматриваемых в этом дополненном докладе: 1) выбор схемы анализа, 2) оценка неопределённости, 3) оценка значений входных параметров, 4) возможно ли (и каким образом) согласовать структуру анализа и входных данных с дополнительным анализом эффективности затрат (cost-effectiveness analysis, CEA), 5) возможно ли оценить выгоды со стороны здоровья и со стороны затрат и 6) в каком виде лучше представить результаты регулятору. Доклад содержит примеры из литературных источников, а также из источников по BIA, размещенных в общем доступе в интернете.

Сопредседатели рабочей группы — *Josephine A. Mauskopf* и *Sean D. Sullivan* — в 2012 году были утверждены Советом Директоров ISPOR. Члены рабочей группы были приглашены сопредседателями для участия по совету и с разрешения Совета Директоров. Члены рабочей группы являются опытными разработчиками или пользователями BIA. Исходя из того, что целью этого доклада была разработка руководства по подготовке и представлению BIA, которое объединило бы разработчиков в области здравоохранения в глобальном плане, в рабочую группу были включены представители регуляторных органов из Северной Америки, Европы, Латинской Америки и Азии.

Рабочая группа провела организационную встречу вовремя 17 Ежегодной Международной встречи ISPOR в Вашингтоне и открытый форум на 18 Ежегодной Международной встрече ISPOR в Нью-Орлеане. До подготовки первого черновика дополненного доклада был произведен целенаправленный поиск по свежим публикациям BIA, для того чтобы оценить методы, встречающиеся в рецензируемых научных журналах. Были рассмотрены шаблоны и методы расчета затрат, руководства по отчетности, выпущенные Национальным Институтом Здоровья и Клинического Совершенствования Великобритании (NICE). Также был изучен опыт Академии Страховой Фармацевтики ISPOR (The Academy of Managed Care Pharmacy Format for Formulary Submission) и доклад Тайваньской Оценки Медицинских Технологий (Taiwan Health Technology Assessments). Были собраны опубликованные местные и национальные руководства, использованные впоследствии как справочные документы. Обсуждение основных принципов и идей в течение последующих месяцев проводилось посредством телефонных звонков и электронной переписки. Разделы доклада были подготовлены различными подразделениями рабочей группы, черновик готового доклада затем был проработан сопредседателями и вновь возвращен рабочей группе для рассмотрения. После этого состоялась очная встреча всех участников рабочей группы для обсуждения черновика и его редактирования перед отправкой на рецензию.

Предварительно была определена главная группа рецензентов, опытных в разработке и использовании BIA. У этой группы были опрошены комментари перед отправкой черновика на более широкое обсуждение. После этого предварительного просмотра новый черновик, исправленный в соответствии с полученными комментариями, был подготовлен ра-

бочей группой для более широкого рецензирования всеми членами ISPOR. В результате конечный доклад отражает вклад как главных, так и всех остальных рецензентов.

Введение

Определение и назначение

Регуляторные организации в области здравоохранения всё чаще обращаются к анализу влияния на бюджет (BIA) наряду с анализом эффективности затрат (CEA) во время работы над созданием списков и документов по возмещению затрат. BIA рассматривает ожидаемые изменения расходов системы здравоохранения после начала использования новых вмешательств (интервенций) при оказании медицинской помощи (рис. 1). BIA может использоваться при планировании бюджета или ресурсов, при этом возможно применение только BIA или BIA совместно с «встроенным» CEA.

Среди пользователей BIA могут быть как лица, управляющие бюджетом и планирующие его, к примеру, руководители государственных или региональных программ здравоохранения, программ индивидуального медицинского страхования, или организации доставки медицинского обеспечения, так и работодатели, напрямую оплачивающие услуги здравоохранения. Им всем необходима наглядно представляемая информация о финансовом влиянии введения в практику и распространения новых интервенций в медицине. Однако требования пользователей могут различаться в зависимости от интересующих их горизонтов времени и ценовых категорий.

Общие положения

Схема анализа для BIA была опубликована *Mauskopf* в 1998 г. [3] Позднее в свет вышли статьи, описывающие методологию BIA [4–8]. Начиная с 1990-х годов, большинство регионов мира, включая Австралию, Северную Америку (Канада и США), Европу (Англия и Уэльс, Испания, Бельгия, Франция, Венгрия, Италия и Польша), Южную Америку (Бразилия и Колумбия), Азию (Южная Корея, Тайвань и Тайланд) и Ближний Восток (Израиль), включили BIA в требования к доказательной базе при составлении национальных и местных формулярных списков и принятии решений о возмещении. Кроме того, были разработаны специфические для каждой страны национальные руководства по построению и представлению BIA. В разделе «Дополнительные материалы / Supplementary Materials» по адресу <http://dx.doi.org/10.1016/j.jval.2013.08.2291> в «Приложении А / Appendix A» можно найти примеры руководств для различных стран с прикре-

пленными URL-ссылками. Основной целью данного доклада было предоставление данных, необходимых для обновления национальных руководств по BIA в каждой стране.

Структура документа

Специалистами признается, что методы выполнения и представления отчетов BIA продолжают развиваться. Этот доклад освещает как области достигнутых соглашений, так и те сферы, где необходимо продолжение развития методологий. Руководство разделено на три основные части: 1) схема/план анализа 2) информационные ресурсы для входящих параметров и других данных 3) формат отчёта.

Рекомендации для схемы/плана анализа

BIA — это способ объединения знаний, доступных на момент принятия решения о создании формулярного списка или покрытия затрат, в виде оценки наиболее вероятных финансовых последствий данного решения для системы здравоохранения. Учитывая локальный характер изучаемых систем и изменяющиеся потребности регуляторных органов, BIA не может дать единую оценку, пригодную для всех регуляторов. Вместо этого, целью BIA является обеспечение валидной схемы анализа — «модели» — которая позволит пользователям, используя свои входные значения, видеть финансовую оценку, соответствующую местной обстановке. Таким образом, результаты BIA должны отражать конкретные сценарии, основанные на специфических локальных допущениях и входных данных, применимых к интересам регулятора.

Этот раздел представляет рекомендации рабочей группы для ключевых элементов аналитической схемы BIA как самостоятельного процесса. Ситуации, в которых BIA дополняется CEA, специально оговорены. Надлежащий дизайн является ключевым шагом в разработке BIA. Дизайн BIA для новой медицинской интервенции должен учитывать локальные особенности системы здравоохранения, возможную ограниченность доступа к медицинским услугам, ожидаемый объем потребления новой интервенции, а также эффективность стандартной терапии и нового медицинского вмешательства. Наш доклад не содержит руководства по моделированию, но освещает ключевые моменты, принимаемые по внимание в процессе планирования BIA, включая выбор методики моделирования. Важно понимать, что какой бы выбор ни был сделан, он должен быть понятным и оправданным и должен обеспечивать максимально простой дизайн, отвечающий нуждам регулятора. Важные для построения схемы анализа BIA аспекты представлены в табл. 1 и описаны в соответствующих разделах.

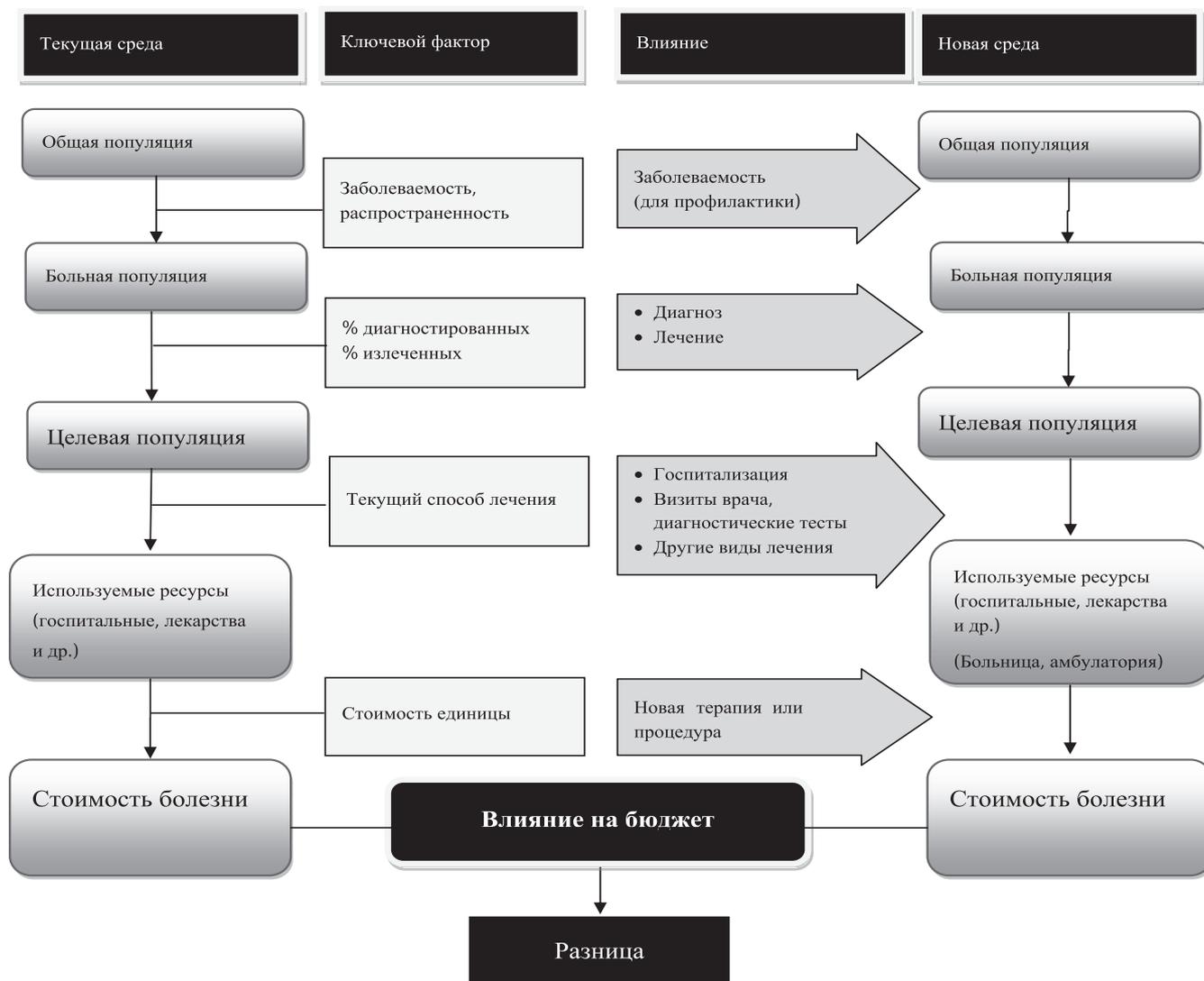


Рис. 1. Схема влияния на бюджет. Адаптирована из Pharmacocon Spanish Res Artic, 2, Brosa M., Gisbert R., Rodríguez Barrios J.M., et al., métodos y aplicaciones del análisis del impacto presupuestario en sanidad, 65–79, 2005. Et al. [2]

Таблица 1

Аспекты, которые необходимо учесть в дизайне анализа влияния на бюджет (ВИА)

- Характеристики системы здравоохранения
- Перспективы
- Использование и стоимость текущих и новых вмешательств
 - ♦ Исследуемая (целевая) популяция
 - ♦ Текущие вмешательства (стандартная терапия)
 - ♦ Внедрение новых медицинских вмешательств и их влияние на рынок
 - ♦ Применение новых медицинских вмешательств вне зарегистрированных показаний
 - ♦ Стоимость стандартной терапии и нового подхода к лечению
- Влияние на другие расходы
 - ♦ Затраты, связанные с состоянием (заболеванием)
 - ♦ Косвенные затраты
- Горизонт моделирования
- Временные зависимости и дисконтирование
- Выбор математической модели
- Оценка неопределенности и анализ альтернативных сценариев
- Проверка достоверности (валидизация)

Особенности системы здравоохранения

Среди особенностей системы здравоохранения важно учесть те из них, которые влияют на бюджет и могут зависеть от правил финансового покрытия (например, повторная госпитализация в 30-дневный срок реабилитации может быть неоплачиваемой, и, таким образом, изменения цен не повлияют на бюджет страховщика, но будут иметь значимое влияние на бюджет медицинского учреждения или поставщика медицинских услуг). Ограниченность доступа к медицинским технологиям также является важной особенностью здравоохранения, которая широко варьирует в зависимости от рассматриваемой системы. Некоторые системы работают по принципу «все или ничего»: либо они обеспечивают полное покрытие затрат, либо вообще никакого; однако в большинстве систем существует множество промежуточных опций, включающих, в том числе, административные или клинические препятствия, ограничения возмещения по субъекту или по видам медицинских технологий, а также различные степени частичной оплаты лечения самими пациентами. Все эти локальные особенности необходимо включить в модель. Это важно сделать не только для новых, но и для уже используемых медицинских вмешательств, в частности, если существует вероятность изменения существующих ограничений для уже используемых вмешательств.

Перспективы

Учитывая цели ВИА, рекомендуется оценивать перспективы с точки зрения спонсора/регулятора. Характеристики регулятора для каждой страны весьма разнятся, при этом его функцию может выполнять как единый орган, покрывающий всю систему здравоохранения через определенные медицинские центры в определенных сферах здравоохранения (например, фармации), так и сам пациент или его семья. ВИА требует гибкости, чтобы давать оценки системам, включающим различные комбинации медицинской помощи, социальных услуг и других расходов при изменяющихся условиях. Например, в одной местности регулятор в области фармации будет иметь отношение только к затратам на лекарственные средства, тогда как в другой области его сфера может включать в себя также бюджет больницы или региона. Обладая гибкой конструкцией, ВИА сможет представить этому регулятору 1) более широкое освещение последствия внедрения новой интервенции, а также 2) влияние на других спонсоров/регуляторов.

Использование и стоимость стандартных и новых интервенций

Целевая популяция. Популяция, включаемая в ВИА, должна состоять из пациентов, соответствующих показаниям применения нового вмешательства

в течение временного горизонта моделирования, учитывая любые имеющиеся ограничения. Определение целевой популяции начинается с оценки количества пациентов, имеющих показания для новой интервенции. Далее следует отразить все запланированные ограничения на использование интервенции, ограничения на покрытие затрат, особенности внедрения (см. следующий раздел) интервенции в практику, включая ее использование не по показаниям, использование интервенции у пациентов, ранее не лечившихся, но желающих начать лечение из-за хороших результатов клинических исследований данной интервенции, большей её убедительности или малого количества заявленных нежелательных явлений (индуцированное рекламой поведение). Также необходимо отразить изменения в терапевтических подходах (например, новая интервенция имеет меньше нежелательных явлений и может быть использована для лечения болезни на более ранних этапах [9]).

Целевая популяция является открытой в том плане, что субъекты могут включаться и исключаться по ходу исследования, в зависимости от их текущего соответствия критериям включения (например, при изменении показаний, появлении ограничений к терапии, исчезновении симптомов, смерти). Если частота входа и выхода из модели стабильны и одинаковы, размер целевой популяции не будет изменяться со временем. Но изменения будут происходить, если новая интервенция повысит длительность нахождения в терапии, замедлит прогрессирование болезни или понизит уровень смертности, не устраняя состояние/заболевание (например, если критерием включения является «умеренная тяжесть заболевания», а новая интервенция замедляет прогрессирование заболевания, то целевая популяция начнет увеличиваться на фоне терапии). Схема анализа должна учитывать влияние этих изменений на размер исследуемой популяции.

Также важно учесть пропорцию, которую занимают представители целевой популяции в определенных подгруппах населения, разделенных по тяжести или стадии заболевания, по наличию и характеру сопутствующих заболеваний, по возрасту, полу, национальности и другим важным характеристикам. Учитываться должны те подгруппы, в которых влияние нового вмешательства на бюджет окажется отличным от общей популяции. Особенно важно включить поправку на подгруппы в схему анализа в том случае, если новая интервенция может изменить пропорции целевых субъектов в подгруппах.

Текущие медицинские вмешательства. Как правило, ВИА сравнивает сценарии, включающие целый ряд вмешательств. Первым следует оценить сценарий использования текущего набора проводимых вмешательств в целевой популяции (стандартная терапия). Одним из вариантов стандартной тера-

пии может являться отсутствие каких-либо вмешательств. Стандартная терапия может включать медицинские вмешательства, которые планируется заменить новой интервенцией. Вмешательства, не имеющие показаний для применения в исследуемой популяции, но применяемые на практике, также могут быть включены в этот сценарий. Конечно, это не должно рассматриваться как поощрение подобного использования медицинских технологий, но как необходимость при описании совокупности проводимых вмешательств. Аналитическая схема должна учитывать вариативность в некоторых частотных и стоимостных параметрах. Если назначение вмешательства требует проведения диагностических тестов, их частоту и стоимость также следует включить в схему анализа в соответствии с принятой местной практикой их применения.

Внедрение новых медицинских вмешательств и рыночные эффекты. Внедрение новой интервенции приводит в движение различные рыночные процессы, включающие замену продуктов и возможное расширение рынка. Будет ли новая интервенция введена и какой эффект она окажет на существующие на данный момент терапевтические подходы во время проведения анализа, неизвестно по определению. Следует предусмотреть три типа возможного внедрения интервенции:

- новая интервенция вытеснит одну или несколько существующих (замещение);
- новая интервенция будет добавлена к списку уже применяемых методов лечения (объединение);
- новая интервенция сможет быть использована в ситуациях, в которых на данный момент не существует эффективного лечения (а есть только поддерживающая терапия), или у пациентов, которые прекратили или отказались от использования доступных методов лечения по причине их непереносимости, возникающего при их применении дискомфорта, развития толерантности или по любым другим причинам (расширение).

Эти виды внедрения могут иметь отношение к различным популяционным подгруппам, и их распределение может изменяться с течением времени. Иными словами, изменения не происходят мгновенно, как только новый продукт попадает на рынок. Поэтому прогнозы таких изменений являются сложным, но очень важным компонентом ВИА.

Изменения имеют различное значение для регулятора. При замещении стоимость новых медицинских вмешательств может быть частично возмещена благодаря сокращению использования стандартных процедур. Возмещение должно производиться не только с учётом замены вмешательства, но и принимая во внимание сопутствующие аспекты, такие

как использование дополнительных лабораторных тестов и купирование нежелательных явлений. Например, применение прямых ингибиторов тромбина не требует фармакокинетического мониторинга, который необходим при использовании варфарина [10], что снижает необходимость в услугах фармакокинетических лабораторий. Объединение приведёт к появлению дополнительных затрат, включая вероятные добавочные стоимости из-за изменений в профиле нежелательного явления при комбинации вмешательств, а также возможную необходимость новых мониторинговых тестов. Расширение увеличит затраты на медицинские вмешательства, но может быть скомпенсировано за счёт лечения ранее неизлечимых пациентов.

Внедрение в практику новых вмешательств и их влияние на стандартный набор применяемых процедур являются ключевыми компонентами ВИА. К сожалению, в большинстве случаев ВИА оперируют слишком малыми объемами известных данных, при этом результаты ВИА могут быть очень чувствительны к используемым допущениям. Анализ должен быть прозрачным по отношению к допущениям, касающимся стандартной терапии и изменений, ожидаемых при внедрении в практику новых методов лечения. Должна ли схема анализа ограничиваться простой оценкой изменения частот во времени или ей следует учитывать более сложные нелинейные функции, в которых внедрение будет зависеть от других факторов (например, от ограниченного доступа к медицинским услугам), зависит от обстоятельств, при этом следует учитывать любую информацию по истории внедрения уже имеющихся схожих вмешательств. Важно, чтобы схема анализа позволяла пользователям просчитать, в том числе, альтернативные варианты результатов внедрения интервенции и их влияние на использование стандартной терапии. Это так называемый анализ структурной неопределенности модели, который имеет особую важность при оценке воздействия на бюджет.

Применение новых медицинских вмешательств без показаний. На практике новые вмешательства могут использоваться для лечения без показаний к их применению. Так как при подобном использовании информация об эффективности и безопасности будет практически полностью отсутствовать, а поощрения подобного применения следует избегать, то включение такого лечения в ВИА не рекомендуется, если только регулятор не указывает этого отдельным запросом.

Стоимость текущего и нового лечения. Стоимость стандартной терапии и нового лечения определяется путём умножения цены, устанавливаемой регулятором за каждое вмешательство, на пропорцию, которую занимает целевая популяция в общей популяции, а также на количество людей в целевой популяции.

Влияние на другие расходы

Затраты на лечение заболевания/состояния.

Введение новых форм терапии может привести к изменению симптомов, длительности, исходов или же темпа прогрессирования болезни и, соответственно, к изменению объёма терапевтических вмешательств. Например, в случае с ВИЧ-инфекцией новый вид терапии, увеличивающий время сохранения у пациента высокого уровня содержания CD4 клеток, скорее всего снизит затраты на лечение ВИЧ-ассоциированных состояний, таких как оппортунистические инфекции. Хотя включение в расчёты затрат на лечение определенных патологических состояний может потребовать использования значительных допущений и ненужного увеличения горизонта моделирования, результат всё же может оказаться важным для некоторых регуляторов. В ВИА, включающем анализ эффективности затрат (СЕА), эти оценки можно получить из самого СЕА. Исходя из вышесказанного, если доступны качественные входные данные, и рассматриваемые изменения действительно имеют влияние на бюджет здравоохранения, то их следует включить в ВИА. И если такое решение будет принято, то модель должна быть составлена таким образом, чтобы пользователь мог увидеть результаты как с учётом влияния этих изменений на бюджет, так и без них.

В некоторых случаях терапия вносит изменения в организацию медицинской помощи (например, дни, проведенные в стационаре, посещения врача на дому), и, таким образом, в пропускную возможность системы. Но, в то же время, эти изменения могут не приводить к изменениям в затратах спонсора/регулятора, так как сама система не всегда может финансово приспособиться к новым условиям в течение рассматриваемого промежутка времени (например, персонал не может быть сокращен или переведен на другие должности). Тем не менее, по-прежнему желательно включить влияние новой интервенции на изменения в медицинском обслуживании, так как это может привести к определенным выводам при планировании здравоохранения и без учета влияния на бюджет.

Косвенные затраты. Влияние новой интервенции на качество и стоимость социального обслуживания, а также любые другие затраты за пределами сферы здравоохранения не должно планомерно учитываться ВИА, так как эти аспекты, как правило, не являются значимыми для регулятора. Исключения возможны, когда ВИА имеет целью информировать частные страховые компании или компании-работодатели. Такие организации проявляют обоснованный интерес к сохранению здоровой и продуктивной рабочей силы, так как они могут окупать затраты на здравоохранение благодаря сохранению работоспособности застрахованных. Другим исключением могут являться системы здравоохранения, осуществляющие свою деятельность на основании налоговых отчислений

в их фонды, в которых потеря доходов вследствие смерти пациента может иметь значительное влияние на финансовую обеспеченность этой системы.

Временной горизонт

ВИА должен быть представлен для заинтересованного спонсора/регулятора в рамках адекватного промежутка времени в соответствии с составляемой сметой и бюджетным периодом (ежемесячным, ежеквартальным или же ежегодным). Наиболее частым временным горизонтом является промежуток в 1 год или 5 лет, при этом результаты должны быть представлены по каждому бюджетному периоду после предполагаемого начала внедрения исследуемой интервенции. План также должен предусматривать и другие временные горизонты. Хотя горизонт моделирования за пределами нескольких лет требует множественных допущений, он может понадобиться, чтобы проиллюстрировать уменьшенные издержки на лечение болезней, которые могут наступить уже после прекращения применения интересующей спонсора терапии (к примеру, лечение хронического гепатита предотвращает развитие цирроза или рака печени, которые являются вероятными событиями в отдаленном будущем после окончания терапии хронического гепатита).

Временные зависимости и дисконтирование

Некоторые параметры ВИА могут изменяться со временем. К этим параметрам относится стоимость используемой валюты (в зависимости от процессов инфляции и дефляции); потребление услуг; возникновение нового вида терапии, внедряемого в схему лечения; изменение цен (например, в связи с истечением срока действия патента на оригинальный препарат); а также изменения в определении/классификации болезни, показаниях и методах терапии. Прогнозирование этих изменений достаточно сложно, но для указанного в ВИА промежутка времени подобные попытки все же должны быть осуществлены, при условии, что допущения прозрачны, обоснованы и соответствуют известным данным.

В СЕА, где ожидаемые эффекты проецируются на будущее, обычной практикой является дисконтирование всех финансовых потоков с вычислением их чистой стоимости на настоящий момент. Этот подход не рекомендуется для ВИА, где интересы спонсора состоят в определении ожидаемого влияния в каждый конкретный момент времени. Таким образом, ВИА должен представлять отчет по финансовым потокам в каждый бюджетный период без дисконтирования. Если регулятор всё-таки захочет рассчитать чистую стоимость проекта в пересчёте на настоящее время, это с лёгкостью может быть сделано, и модель может предоставлять такую возможность, хотя обычно это не требуется.

Выбор математической модели

Математическая модель составления сметы для ВИА может быть выполнена в форме простого калькулятора затрат на базе Excel (см. ссылки [11—14]), а также образцы формирования стоимости, предоставляемые NICE, которые доступны по www.nice.org.uk). Такой подход является предпочтительным, так как он хорошо понятен регулятору. В условиях, когда изменения в размере целевой популяции, структуре тяжести заболевания или схемах лечения не могут быть убедительно отражены с помощью Excel-калькулятора, для сравнения стоимостей текущей и планируемой схем лечения для целевой популяции могут быть использованы когортные или субъектно-ориентированные модели, должным образом учитывающая количество и качество входящих и покидающих популяцию пациентов (см. ссылки [15—22]). В любом случае эти модели должны соответствовать методам, описанным в руководстве Рабочей Группы ISPOR-SMDM о надлежащей практике моделирования, и вся документация по моделям должна быть доступна регулятору [23].

Неопределенность и анализ альтернативных сценариев

Для ВИА значима неопределенность двух типов: неопределенность параметров, рассчитанных на основе входных данных, и неопределенность структуры модели, представленная допущениями, принятыми при создании ВИА [24]. Примеры неопределенности параметров включают в себя оценку эффективности стандартной и новой терапии, а примером структурной неопределенности являются ожидаемые изменения в схемах лечения при внедрении новой терапии или ожидаемые ограничения к её внедрению. Так как информация по многим параметрам ограничена, большая часть неопределенности параметров в ВИА не может быть достоверно рассчитана, и таким образом стандартные подходы, как, например, однофакторный и вероятностный анализы чувствительности, не могут быть проведены в полной мере. Большая же часть неопределенности является структурной, и параметризовать её нелегко. По этим причинам следует создать правдоподобные альтернативные сценарии, изменяя определённые входные параметры и структурные допущения.

Проверка достоверности (валидизация)

Математическая модель и входные данные, используемые при ВИА, должны быть в достаточной мере достоверными (валидными), чтобы обеспечить регулятора надёжной информацией к моменту при-

нятия им решения. При проведении ВИА должны производиться два стандартных этапа проверки достоверности: 1) определить внешнюю достоверность посредством соглашения с сотрудниками регуляторного органа относительно адекватности математической модели и других аспектов исследования (например, эти люди могут дать консультацию по поводу горизонта моделирования или возможных ограничений доступности исследуемой интервенции); и 2) контроль правильности работы калькулятора, включая все его формулы [25]. В дополнение к этому, по возможности, следует сравнить адекватность оценок, выдаваемых моделью для первого года стандартной терапии с уже имеющимися реальными данными. А после внедрения исследуемой интервенции необходимо выполнить обсервационное исследование и сравнить его результаты с оценками ВИА. Конечно, после принятия решения это уже будет неважно, но обеспечит уверенность в адекватности использованного подхода при построении следующих моделей.

Рекомендации по источникам входных данных

Когда схема анализа разработана, она заполняется необходимыми входными данными. Полезность ВИА частично зависит от качества, точности и применимости данных. Чтобы обеспечить достоверную оценку влияния на бюджет, данные должны поступать из лучших доступных источников и в достаточной мере обеспечивать прозрачность и воспроизводимость. В этом разделе представлены рекомендации рабочей группы по возможным источникам данных для каждого из следующих пяти элементов ВИА:

- Размер и характеристики целевой группы;
- Совокупность медицинских вмешательств, включая и исключая новое вмешательство;
- Стоимость стандартной и новой терапии;
- Использование и стоимость другого медицинского обслуживания, обусловленного состоянием пациента; и
- Разброс значений и результаты альтернативных сценариев для учёта степени неопределенности.

Основные рекомендации

Разработчик ВИА должен быть знаком со специфическими требованиями регулятора к информации. Наиболее важно, чтобы входные данные были значимыми для регулятора. Например, данные по одной стране могут недостоверно отражать ситуацию для другой страны. Источники информации должны соответствовать изучаемому вопросу, и их надёжность должна быть оценена. Табл. 2 обобщает примеры источников информации, которые могут быть использованы при ВИА.

Примеры источников данных для внесения в анализ влияния на бюджет

- Информация о реальных затратах, полученная из регистров или баз данных и отражающая интерес регулятора.
- Информация, полученная из РКИ, специфичная для интересующей регулятора популяции или распространяющаяся на неё.
- Информация по внедрению интервенции, её использованию, а также приверженности к ней пациентов из международных источников по схожим популяциям и со схожими схемами лечения.
- Данные маркетинговых исследований для выявления распределения терапии сравнения на рынках, практик раннего использования альтернативных вмешательств и основных трендов в общепринятых схемах лечения.
- Мнение экспертов и проведение опросов среди специалистов.

Размер и характеристики целевой популяции

Для ВИА необходима оценка размера целевой популяции и распределения любых переменных, способных оказать влияние на бюджет. Предпочтительным подходом будет являться получение оценки размера и характеристик целевой группы напрямую от регулятора. Так как эта информация обычно не является общедоступной, то для получения оценок для ВИА можно использовать альтернативные методы. Если данные регулятора недоступны, можно использовать национальные или местные базы данных по заболеваемости и распространённости интересующего состояния/заболевания с поправкой на пол, возраст, национальность или же группу риска и другие важные переменные. Например, чтобы оценить численность популяции людей с ВИЧ-инфекцией для местной программы медицинского обслуживания, существующая оценка национальной распространённости инфекции может потребовать некоторой коррекции, если в регионе процент людей с высоким риском значимо отличается от такового по национальным данным.

Когда размер популяции с интересующим исследователя состоянием/заболеванием будет оценён, необходимо оценить долю людей с диагностированным заболеванием, долю людей, получающих терапию, а также распределение степеней тяжести заболевания. Всё это подразумевает использование строгих критериев включения для оценки количества людей, соответствующих критериям целевой популяции. Эти оценки должны быть получены от регулятора, но могут быть получены и из других источников. Например, для оценки распределения стадий заболевания или уровней тяжести болезни в целевой популяции могут использоваться исторические работы, отслеживающие тренды распространения хронических заболеваний на протяжении длительных отрезков времени.

Исследователю важно оценить не только текущий размер целевой популяции и структуру тяжести заболевания, но и то, как эти оценки будут изменяться на протяжении моделируемого отрезка времени без или с введением нового вида вмешательства. Примерами подобных изменений в размере популяции и/или структуре тяжести заболевания под действи-

ем лечения могут служить увеличение ожидаемой продолжительности жизни больных раком в терминальной стадии или снижение тяжести заболевания в связи с повышением уровня CD4-лимфоцитов ВИЧ-инфицированных. Эти изменения могут влиять на бюджет, и, по возможности, должны быть учтены в модели на основании результатов РКИ или из регистров. Кроме того, на основе имеющихся литературных данных или мнения специалистов должна быть отмечена любая диагностическая интервенция, которая может способствовать увеличению целевой популяции за счёт лучшего выявления заболевших или более точного определения степени тяжести болезни.

Одно важное соображение, касающееся хронических состояний, заключается в том, что влияние на бюджет новой терапии в течение первых нескольких лет после её внедрения будет иметь эффект намерстывания. Например, если новое лекарство показано тем, кто ранее безуспешно получал специфическую терапию хронического заболевания, то в каждом моделируемом году в целевой популяции придётся учитывать две группы пациентов: тех, кто только что закончил безуспешную стандартную терапию и перешёл на новое лечение, и тех, кто закончил стандартную терапию раньше (год или более года назад) и уже изменил своё лечение либо на исследуемое, либо на какое-то ещё, которое снова может оказаться менее эффективным, чем исследуемое (группа «намерстывания»). Чтобы эффективно включать в модель ВИА новых пациентов и пациентов из группы «намерстывания», от регулятора или из релевантных эпидемиологических исследований должна быть получена информация по размеру каждой группы пациентов, соответствующей критериям включения.

Стандартные и новые схемы лечения

Рекомендованным источником информации о терапии в настоящее время являются данные самого регулятора. Если эти данные недоступны, то можно также использовать опубликованную информацию по текущим схемам терапии из различных стандартов, финансовых отчётов страховых компаний, локальных маркетинговых исследований или любых других вторичных источников. Будучи отнесёнными к «текущим», эти схемы лечения, тем не менее, могут изме-

няться со временем даже при отсутствии внедрения новых интервенций, и это также должно быть отражено в ВИА. Оценка изменений в схеме лечения должна основываться на произошедших ранее изменениях для похожих медицинских технологий, исследовании рынка или на мнении независимых экспертов. Примером могут служить изменения, обусловленные окончанием действия патента на текущую терапию в пределах горизонта моделирования, что, предположительно, приведёт к её более широкому использованию. Если же существует множество медицинских вмешательств, но широко используются лишь немногие из них, то при моделировании стандартной терапии можно ограничиться только ими, но при этом в модели следует оставить место для самостоятельного добавления регулятором других, более редко используемых стандартных методик, на которые может оказать влияние новая интервенция.

Новая схема лечения зависит как от скорости внедрения в практику новой интервенции, так и от меры замещения или расширения с её помощью методов стандартной терапии. Скорость внедрения в практику нового вмешательства изменяется по мере того, как врач и пациент знакомятся с ним. Существует несколько рекомендованных способов оценки уровня внедрения новых медицинских вмешательств в схемы лечения: 1) использование данных из других областей медицины, в которых это вмешательство уже используется; 2) использование оценки ожидаемого удельного веса производителя на рынке, и 3) экстраполирование опыта внедрения похожих медицинских технологий. С помощью всех трёх подходов необходимо определить, будет ли новый вид вмешательства добавлен к стандартной терапии или он заменит её частично или целиком. Исследование рынка, оценка производителя и мнение независимых экспертов являются рекомендованными источниками информации для оценки уменьшения использования стандартной терапии, когда новое вмешательство станет доступным.

Стоимости стандартного и нового терапевтических подходов

Оценка затрат должна быть основана как на фактической себестоимости вмешательства с позиции регулятора (включая все возможные скидки, частичное возмещение затрат, а также любые другие применимые коррективы), так и на дополнительных затратах на необходимую диагностику. Если данные по фактической себестоимости недоступны, аналитик должен использовать общедоступные данные, такие как стоимость оптовой закупки, цены по прейскурантам и зарегистрированные цены. Анализ должен быть проведён таким образом, чтобы все скидки, удерживаемые страхователем суммы, а также долевые платежи могли быть вычтены из этих стоимостей, позволяя

отразить реальные суммы, требуемые от спонсора. В случае с оборудованием и расходными материалами стоит использовать либо оптовые, либо оговоренные в системе здравоохранения цены; в случае с процедурами и диагностикой помимо цен на оборудование и расходные материалы необходимо также учитывать трудовые выплаты персоналу. Стоимость процедуры также должна отражать все затраты, связанные с управлением и мониторингом. Эти затраты оцениваются в соответствии с местными ценами на медицинское обслуживание. Если последние неизвестны для данной местности, то для оценки необходимого медицинского обслуживания следует использовать инструкции по применению препаратов или публикации, описывающие требования к управлению и мониторингу при проведении вмешательства.

Стоимость коррекции нежелательных явлений и осложнений также должна быть включена отдельным пунктом в общую стоимость стандартной или исследуемой терапии. Частота нежелательных явлений вытекает из инструкций по применению или других публикаций для каждого из включаемых в модель вмешательств, что в итоге даст ожидаемую оценку затрат на их коррекцию. Если опубликованные исследования по нежелательным явлениям недоступны, то алгоритмы лечения должны разрабатываться с привлечением врачей, занимающихся лечением данной патологии.

Составление сметы по новому комплексу медицинских вмешательств происходит так же, как при расчете стоимости стандартной терапии, за исключением новых вмешательств, которые ещё не появились на рынке. Стоимость нового вмешательства оценивается с помощью допущений, если оно всё ещё недоступно для популяции на данном этапе, и потому для него неизвестен фактический эффективный режим дозирования. Оценка дозирования должна быть прозрачной и обоснованной, если это возможно, на опыте дозирования аналогичных лекарств. Стоимость введения нового лекарства, мониторинга и лечения возникающих нежелательных явлений при его использовании должна быть оценена аналогично анализу стоимости стандартной терапии с использованием необходимых обоснований. При обращении к независимой экспертной оценке, необходимо принять все меры для того, чтобы вопросы в опросниках были сформулированы правильным образом, а выбор специалистов максимально соответствовал целям и задачам исследования.

Применение и стоимость других видов терапии, ассоциированных с состоянием пациента

Если в расчёты необходимо включить затраты, связанные с заболеванием, то следует предпринять следующие шаги: оценить изменения показателей здоровья; выявить соответствующие ресурсы,

использование которых может меняться; оценить масштаб изменений и их стоимость. В ВИА выявление изменений должно соответствовать интересам регулятора. Чтобы оценить изменения показателей здоровья, необходимо использовать данные клинических исследований. Встроенный СЕА может использоваться в качестве дополнительного источника информации. Частота и объём пользования медицинскими услугами в течение моделируемого горизонта времени должны быть оценены с помощью локальных источников. В отсутствие этих данных к рассмотрению могут быть приняты консультации специалистов, регулярно занимающихся лечением пациентов с интересующей исследователя патологией. Себестоимость каждой медицинской услуги, входящей в интересы регулятора, должна рассчитываться с учётом возможных изменений в частоте её использования для получения адекватного влияния всех изменений, вызванных новой интервенцией, на бюджет лечения. Такие дополнительные расходы могут быть весьма важными. При этом может оказаться очень сложно произвести такую оценку в некоторых областях медицинских услуг в связи с особенностями системы оплаты. В этом случае, если рассчитать альтернативные издержки не представляется возможным, можно использовать другие подходы к расчётам.

Диапазоны и альтернативные значения для анализа неопределённости и различных сценариев

Область значений, используемая в анализе неопределённости, должна быть получена от регулятора. Диапазоны по умолчанию следует получать из обзоров опубликованных исследований или из консультаций врачей, занимающихся лечением пациентов с интересующим исследователя состоянием. Использование произвольных диапазонов, таких как $\pm 20\%$ или $\pm 50\%$ не рекомендуется, так как они не отражают фактическую неопределённость параметров, особенно с точки зрения регулятора [24].

ВИА также включает много параметров, значения которых будут зависеть от регулятора, например, возрастной и половой состав популяции, учитываемой в плане медицинского страхования или системе здравоохранения; схемы терапии и ограничения их использования, применяемые в системе здравоохранения или медицинского страхования; заболеваемость и распространённость болезни; а также стоимость лекарств и медицинского обслуживания населения. Конкретную информацию спонсора при этом стоит использовать для рассмотрения альтернативных правдоподобных сценариев, варьируя значения входных параметров.

Если в анализе сценария учитывается влияние приверженности пациентов к терапии, то на основе исследования баз данных и проспективных иссле-

дований, соответствующих запросам регулятора, должны быть сделаны необходимые допущения. ВИА также должен оценивать влияние приверженности к лечению на его эффективность и безопасность. Если не существует опубликованных данных о взаимосвязи приверженности к лечению с результатами лечения, то возможно использовать альтернативные источники информации, например, данные о фармакокинетике или фармакодинамике препарата или же мнение экспертов.

Рекомендации по форме отчёта

Последний раздел руководства представляет рекомендуемый формат изложения для ВИА. Его целью является обеспечить логичность и прозрачность отчёта.

Введение в отчёте

Введение должно включать цели и задачи анализа, краткое резюме по актуальной эпидемиологической, клинической и экономической информации, касающейся целевой популяции и интересующего заболевания.

Задачи анализа. Задачи ВИА должны быть ясно обозначены и связаны с перспективой заказчика.

Эпидемиология и лечение заболевания. Вступление должно содержать информацию по показателям заболеваемости и распространённости состояния, его тяжести и склонности к прогрессированию, долю не диагностированных и нелеченых случаев, а также информацию по факторам риска, которые необходимо учитывать при оценке влияния на бюджет.

Клиническое влияние. Клинические данные должны включать в себя краткое описание подходящей под критерии включения популяции и существующие варианты вмешательств, их эффективность и безопасность, которые будут являться релевантными для дизайна ВИА.

Экономическое влияние. Информация об экономическом влиянии должна включать в себя краткое описание предыдущих ВИА для интересующего состояния при исследовании других медицинских вмешательств и специфических для данного состояния схем лечения, а также исследования стоимости медицинского ухода для этого состояния.

Дизайн и методы исследования

Доклад должен определять дизайн ВИА (обычно это калькулятор стоимости на базе Excel, но может быть и когортная или субъектно-ориентированная модель). Следующие характеристики схемы анализа должны быть описаны со ссылкой на источник:

Группа больных. Этот раздел должен чётко обозначить целевую популяцию для применения нового вмешательства.

Набор вмешательств. Этот раздел должен содержать подробное описание использования и характеристик каждого из вмешательств в составе стандартной терапии и ожидаемого набора вмешательств в исследуемой терапии. Вмешательства имеют различную стоимость в зависимости от системы здравоохранения, и эти предполагаемые цены должны быть отражены в таблицах [26]. Некоторые пациенты могут получать комбинированную терапию, что должно быть отражено в отдельных таблицах.

Также должны быть предоставлены важнейшие характеристики каждого вмешательства в схеме лечения: показания, дозы, эффективность, нежелательные явления, комплаентность.

Горизонт моделирования. Горизонт моделирования должен быть обозначен и обоснован.

Перспектива. Этот раздел должен ясно определять перспективы ВИА, то есть позицию, с точки зрения которой будет произведён расчёт (с позиции пациента, клиники, страховой компании или регулятора), категории включенных в ВИА стоимостей и целевую популяцию. Всё это будет зависеть от того, для какой структуры или организации выполняется анализ.

Описание схемы анализа. Этот раздел должен содержать подробное описание структуры калькулятора стоимости, когортной или субъектно-ориентированной модели. Следует включить в отчёт графическое представление модели, например, схему анализа.

Входные данные. Должны быть представлены все входные значения, в том числе и для альтернативных сценариев, используемые в анализе. Уровень детализации должен быть достаточен для того, чтобы читатель мог воспроизвести все расчёты по модели самостоятельно.

Источники информации. Источники входных данных должны быть детально описаны. Любые изменения или пересчёты показателей должны быть описаны достаточно детально, чтобы дать возможность их воспроизвести. Сильные и слабые стороны, ограничения, возможное направление и размер ошибки — всё должно быть отмечено. Должны быть обсуждены критерии выбора исследований и баз данных, являющихся источниками входной информации.

Сбор информации. Методы и процесс сбора первичных данных, а также их адаптации для модели должны быть описаны и объяснены. Формы сбора данных или вопросники должны быть включены в приложение к отчёту.

Анализ. Должно быть предоставлено описание расчётов, использованных при ВИА. Выбор всех представленных в результатах сценариев должен быть документирован и обоснован.

Неопределённость. Методы анализа неопределённости должны быть описаны и обоснованы.

Результаты

Результаты должны быть представлены детально с целью обеспечения гибкости при их использовании регулятором и другими заинтересованными лицами, с которыми он будет взаимодействовать. Влияние на бюджет должно быть представлено для каждого бюджетного периода в течение анализируемого промежутка времени. Также должна быть предоставлена информация о затратах и об использованных ресурсах за данный бюджетный период. Оценки использования ресурсов должны быть отражены в таблице, показывающей изменения в их использовании для каждого подотчётного периода ВИА. Оценки использования ресурсов должны быть группированы по вмешательствам и по нежелательным явлениям. Отдельная таблица должна показывать общие затраты и затраты по отдельным направлениям (лекарства, посещения врача, амбулаторное и стационарное лечение, уход за пациентом на дому) для каждого отчётного периода. Также могут быть предоставлены изменения в ежегодных показателях результатов лечения. Всё это позволит обеспечить ясность для регулятора, когда значительное влияние на бюджет будет сопровождаться не менее значительным положительным влиянием на здоровье пациентов. Результаты анализа неопределённости и анализа альтернативных сценариев должны быть представлены в виде диаграмм и таблиц.

Выводы и ограничения

Необходимо обозначить основные выводы по основным результатам ВИА. Следует предоставить сведения об основных ограничениях в отношении ключевых вопросов, а именно — аспектов дизайна, включающих в себя использование в отсутствие показаний, допущения о несоблюдении рекомендаций, а также полноту и качество информации об исходных ресурсах и данных.

Использование графиков и таблиц

Рисунки и схемы могут быть полезны для пользователей, желающих скопировать их с целью включения в свою внутреннюю отчётность. Для этого рекомендуется использовать следующие инструменты:

Графическое изображение структуры анализа. Графическое представление структуры анализа облегчает для спонсора понимание механизма подсчётов изменения затрат и ресурсов. В раздел описания схемы анализа рекомендуется включить структурные диаграммы или другие визуальные изображения калькулятора затрат, когортной или субъектно-ориентированной моделей.

Таблица допущений. Все основные допущения должны быть представлены в табличной форме.

Таблицы входных параметров. Все значения входных параметров и их источники должны быть представлены в виде таблиц.

Таблицы полученных результатов. Все полученные результаты должны быть представлены в табличной и/или графической форме.

Схематическое представление анализа неопределённости. Торнадо-диаграммы, отображающие эффект изменения каждого из входящих в модель параметров в заданных диапазонах, являются простым способом для отображения ключевых факторов, влияющих на результат ВИА, и должны быть включены в раздел, посвящённый результатам анализа сценариев.

Приложения и ссылки

Поощряется добавление релевантных приложений к основному отчёту. Приложения могут включать в себя стратегии поиска литературы, резюмирование фактических данных, промежуточные результаты, а также имена и адреса участвующих экспертов и исследователей.

Отчёт ВИА с включенным в него СЕА

За последнее время было опубликовано несколько статей на тему дизайна, методов и полученных результатов при комплексном использовании ВИА и СЕА [16, 27]. Хотя они и взаимосвязаны, дизайны этих исследований требуют различных структурных решений и использования ряда различных и/или дополнительных параметров. Зачастую они рассчитываются исходя из различных перспектив и используют различные временные диапазоны. Сообщая о совместно проведённых ВИА и СЕА в одной работе, аналитик должен докладывать обо всех элементах обоих анализов. В дальнейшем мы советуем докладывать о ВИА, придерживаясь описанного выше формата, а о СЕА — в соответствии с форматом объединённых стандартов доклада об оценке экономики здравоохранения Consolidated Health Economic Evaluation Reporting Standards (CHEERS) [28]. Если журнал или доклад имеют ограничения объёма, то к статье необходимо предусмотреть приложение, чтобы предоставить необходимые детали.

Компьютерная программа влияния на бюджет

Так как ВИА требует достаточной гибкости для обеспечения оценок разным заказчиком, очень важно, чтобы Excel-калькулятор, когортная или субъектно-ориентированная модели позволяли пользователям вносить их собственные значения затрат. Программа также должна позволять пользователям с лёгкостью восстанавливать изначальные стандартные параметры. Программа должна содержать про-

стые в использовании таблицы на общедоступной платформе программного обеспечения. Все входящие параметры должны отображаться в виде одной или нескольких рабочих таблиц. Будет полезным обозначение граф входных параметров различными цветами в зависимости от того, может ли пользователь вводить в них свои данные, или нет. Выходные данные должны отображаться в одной или нескольких рабочих таблицах в логической манере, которая резюмирует все результаты, полученные пользователем. Полезно использовать графическое изображение выходных данных. Иногда спонсор может быть заинтересован в анализе более чем одного вопроса. В таких случаях интерфейс пользователя должен быть спроектирован таким образом, что пользователь мог с лёгкостью переключаться между различными анализируемыми вопросами. Должны быть предоставлены образцы анализа неопределённости и альтернативных сценариев. Программа должна позволять пользователю производить анализ сценариев и анализ неопределённости в соответствии с этими образцами. Первый лист в Excel должен включать оглавление структуры модели, описание допущений и инструкцию к использованию модели. Все источники, вычисления и допущения, применимые к входным данным, должны быть отражены, должен быть приведён список всех использованных ссылок. Все расчёты должны быть доступны пользователю и представлены в полном объёме. Также должно быть включено подробное руководство для пользователей.

Заключение

ВИА важны для экономической оценки новых медицинских вмешательств. Этот обновленный доклад рабочей группы ISPOR включает руководство по разработке схемы анализа для ВИА, даёт рекомендации по используемым источникам данных, описывает единый формат отчёта, который будет способствовать последовательности и прозрачности анализа, а также описывает требования к оформлению компьютерной программы для ВИА. Следование предложенным принципам надлежащей исследовательской практики не обязательно имеет целью вытеснить руководства по проведению ВИА, имеющиеся в разных странах, но должно помочь поддержать и улучшить местные рекомендации или же послужить отправной точкой для написания новых.

Благодарности

Специалисты, входящие в состав основной группы рецензентов, и предоставившие ценные указания и комментарии по первому черновику доклада рабочей группы: *Lieven Annemans*, PhD, *C. Daniel Mullins*, PhD, и *Bong-Min Yang*, PhD. Авторы также выражают благодарность 84 членам ISPOR из 24 стран за под-

робные комментарии к более ранней версии доклада, а также *Maria Swift* из офиса ISPOR за блестящую техническую и административную поддержку на всех этапах деятельности рабочей группы.

Источник финансовой поддержки: Авторы не по-

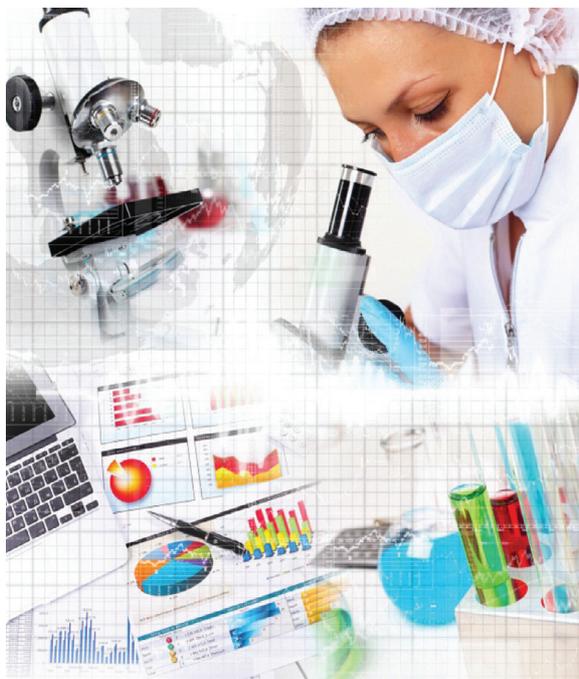
лучали финансовой поддержки за их участие в рабочей группе. Все авторы добровольно пожертвовали своим временем для обсуждения, исследования и написания доклада.

Литература

1. *Mauskopf J.A., Sullivan S.D., Annemans L., et al.* Principles of good practice for budget impact analysis: report of the ISPOR Task Force on Good Research Practices — Budget Impact Analysis. *Value Health* 2007;10:336—47.
2. *Brosa M., Gisbert R., Rodríguez Barrios J.M., Principios Soto J.* métodos y aplicaciones del análisis del impacto presupuestario en sanidad. *Pharmacoecon Spanish Res Artic* 2005;2:65—79.
3. *Mauskopf J.* Prevalence-based economic evaluation. *Value Health* 1998;1:251—9.
4. *Trueman P., Drummond M., Hutton J.* Developing guidance for budget impact analysis. *Pharmacoeconomics* 2001;19:609—21.
5. *Orlewska E., Mierzejewski P.* Proposal of Polish guidelines for conducting financial analysis and their comparison to existing guidance on budget impact in other countries. *Value Health* 2004;7:1—10.
6. *Nuijten M., Mittendorf T., Persson U.* Practical issues in handling data input and uncertainty in a budget impact analysis. *Eur J Health Econ* 2011;12:231—41.
7. *Garattini L., van de Vooren K.* Budget impact analysis in economic evaluation: a proposal for a clearer definition. *Eur J Health Econ* 2011;12:499—502.
8. *Garay O.U., Caporale J.E., Pichón-Riviere A., et al.* Budgetary impact analysis in health: update with a model using generic approach. *Rev Peru Med Exp Salud Publica* 2011;28:540—7.
9. Department of Health and Human Services Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Available from: <http://aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf>. [Accessed March 27, 2012].
10. *Goel R., Svivathsan K.* Newer oral anticoagulant agents: a new era in medicine. *Curr Cardiol Rev* 2012;8:158—65.
11. *Chang J., Sung J.* Health plan budget impact analysis for pimecrolimus. *J Manag Care Pharm* 2005;11:66—73.
12. *Danese M.D., Reyes C., Northridge K., et al.* Budget impact model of adding erlotinib to a regimen of gemcitabine for the treatment of locally advanced, nonresectable or metastatic pancreatic cancer. *Clin Ther* 2008;30:775—84.
13. *Dee A., Hutchinson M., De La Harpe D.* A budget impact analysis of natalizumab use in Ireland. *Ir J Med Sci* 2012;181:199—204.
14. *Smith D.G., Cerulli A., Frech F.H.* Use of valsartan for the treatment of heart-failure patients not receiving ACE inhibitors: a budget impact analysis. *Clin Ther* 2005;27:951—9.
15. *Mauskopf J.* Meeting the NICE requirements: a Markov model approach. *Value Health* 2000;3:287—93.
16. *Marchetti M., Caruggi M., Colombo G.* Cost utility and budget impact of third-generation aromatase inhibitors for advanced breast cancer: a literature-based model analysis of costs in the Italian National Health Service. *Clin Ther* 2004;26:1546—61.
17. *Caro J.J., Huybrechts K.F., Xenakis J.G., et al.* Budgetary impact of treating acute bipolar mania in hospitalized patients with quetiapine: an economic analysis of clinical trials. *Curr Med Res Opin* 2006;22:2233—42.
18. *Mar J., Sainz-Ezkerra M., Miranda-Serrano E.* Calculation of prevalence with Markov models: budget impact analysis of thrombolysis for stroke. *Med Decis Making* 2008;28:481—90.
19. *Purmonen T.T., Auvinen P.K., Martikainen J.A.* Budget impact analysis of trastuzumab in early breast cancer: a hospital district perspective. *Int J Technol Assess Health Care* 2010;26:163—9.
20. *Martin E.G., Paltiel A.D., Walensky R.P., Schackman B.R.* Expanded HIV screening in the U.S.: what will it cost government discretionary and entitlement programs? A budget impact analysis. *Value Health* 2010;13:893—902.
21. *Mar J., Arrospide A., Comas M.* Budget impact analysis of thrombolysis for stroke in Spain: a discrete event simulation model. *Value Health* 2010;13:69—76.
22. *Colin X., Lafuma A., Costagliola D., et al.* Modelling the budget impact of darunavir in the treatment of highly treatment-experienced, HIV-infected adults in France. *Pharmacoeconomics* 2010;28(Suppl. 1):183—97.
23. *Caro J.J., Briggs A.H., Siebert U., et al.* Modeling good research practices — overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—1. *Value Health* 2012;15:796—803. [24]
24. *Briggs A.H., Weinstein M.C., Fenwick E., et al.* Model parameter estimation and uncertainty analysis: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—6. *Value Health* 2012;15:835—42.
25. *Eddy D.M., Hollingworth W., Caro J.J., et al.* ISPOR-SMDM Modeling Good Research Practice Task Force. Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force—7. *Value Health* 2012;15:843—50.
26. *Arlandis S., Castro D., Errando C., et al.* Cost-effectiveness of sacral neuromodulation compared to botulinum neurotoxin or continued medical management in refractory overactive bladder. *Value Health* 2011;14:219—28.
27. *Chiao E., Meyer K.* Cost effectiveness and budget impact of natalizumab in patients with relapsing multiple sclerosis. *Curr Med Res Opin* 2009;25:1445—54.
28. *Husereau D., Drummond M., Petrou S., et al.* Consolidated Health Economic Evaluation Reporting Standards (CHEERS)—explanation and elaboration: a report of the ISPOR Health Economic Evaluation Publication Guidelines Good Reporting Practices Task Force. *Value Health* 2013;16:231—50.



Маркет Аксес Солюшенс — это комплекс услуг по продвижению лекарственных средств, медицинского оборудования и медицинских технологий



Миссия:

Выявляя и продвигая новые медицинские технологии, повышать качество оказания медицинской помощи и эффективность системы здравоохранения

Фокус:

- Формирование уникального восприятия ценности продукта
- Позиционирование продукта на рынке
- Определение целевых областей применения
- Эффективное взаимодействие PR

Наши услуги:

- Анализ рынка и системы здравоохранения
- Маркетинговые исследования
- Разработка стратегии
- Регуляторные задачи
- Клинические исследования
- Оценка медицинских технологий
- Определение ценовой политики и целевого финансирования
- Построение эффективных бизнес-процессов для фармацевтических компаний



Мы всегда рады сотрудничеству:

☎ +7 (495) 664-32-70

✉ info@marketaccess.ru

🌐 www.Market-Access-Solutions.ru

