

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_{st} 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 рационированием 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 существует 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, используемых 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 используемых 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.I., 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