

## Болезнь-специфичный опросник по исходам, сообщаемым пациентами с артериальной гипертензией. Часть II: валидационное исследование и отбор вопросов

Ионов М. В.<sup>1,2</sup>, Звартау Н. Э.<sup>1,2</sup>, Дубинина Е. А.<sup>3,4</sup>, Хромов-Борисов Н. Н.<sup>1</sup>, Конради А. О.<sup>1,2</sup>

**Цель.** Повышение качества жизни, связанного со здоровьем (КЖ) — один из принципов ценностно-ориентированной медицины. КЖ оценивается показателями исходов, сообщаемых пациентами (ПИСП), в случае артериальной гипертензии (АГ) используются лишь базовые опросники. Экспертами была создана первичная версия АГ-специфичного ПИСП. Цель второго этапа — проведение крупного валидационного исследования и статистически-обоснованный отбор вопросов.

**Материал и методы.** Анкетирование пациентов с АГ 1-3 степеней тяжести и здоровых добровольцев проводилось в многопрофильном центре. Критерии включения: возраст >18 лет, способность самостоятельно прочесть и ответить на вопросы, отсутствие значимой патологии, требующей госпитализации. Проведена селекция вопросов в соответствии с принципами классической теории тестов (КТТ) и современной теории тестирования (СТТ). Критерии КТТ: оценка чувствительности (стандартное отклонение и коэффициент вариации, соответствующие доверительные интервалы), репрезентативности (коэффициент корреляции “элемент-итог” Пирсона), внутренней согласованности (коэффициент  $\alpha$  Кронбаха). Критерии СТТ: оценка сложности вопроса и коэффициент дискриминации  $\alpha$ . Каждый вопрос оценивался по 8 критериям. В случае несоответствия  $\geq 4$  критериям обсуждалось удаление вопроса из общего пула. Экспертами определялась практическая важность элементов.

**Результаты.** В анкетировании участвовало 430 человек. Возвращено 407 (94,7%) заполненных ПИСП (от 359 пациентов с АГ,  $62,3 \pm 1,7$  лет; от 48 условно здоровых добровольцев,  $38,8 \pm 10,5$  лет), среднее время заполнения —  $24 \pm 4,2$  мин. Из 163 вопросов 27 отвечали всем критериям, 3 вопроса — ни одному. Из 36 вопросов АГ-части 11 отвечали  $\geq 5$  критериям; в общей части — 87 вопросов (33 — в домене физиологии (PHY), 35 — в психологическом (PSY), 8 — в социальном (SOC), 11 — в сфере лечения (THER)). Симметричное распределение критериев было у 25 вопросов, из которых 11 оценены экспертами и оставлены. Для 40 вопросов было зарегистрировано <4 критериев соответствия, 9 из них оставлены после экспертной оценки. Итоговый вариант версии ПИСП содержал 80 вопросов (19 вопросов в сфере PHY, 22 — PSY, 6 — SOC, 13 — THER, 20 АГ-специфичных).

**Заключение.** Методики КТТ и СТТ позволили сократить объем ПИСП без потери смысловой насыщенности и необходимости реорганизации понятийной структуры. Следующим этапом запланирована валидация ПИСП.

**Конфликт интересов:** не заявлен.

**Финансирование.** Исследование выполнено за счет гранта Российского научного фонда (проект № 17-15- 01177).

<sup>1</sup>ФГБУ Национальный Медицинский Исследовательский Центр им. В. А. Алмазова Минздрава России, Санкт-Петербург; <sup>2</sup>ФГАОУ ВО Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики (Университет ИТМО), Санкт-Петербург; <sup>3</sup>Российский государственный педагогический университет им. А. И. Герцена, Санкт-Петербург; <sup>4</sup>Национальный медицинский исследовательский центр психиатрии и неврологии им. В. М. Бехтерева, Санкт-Петербург, Россия.

Ионов М. В. — аспирант, м.н.с. научно-исследовательской лаборатории патогенеза и терапии артериальной гипертензии научно-исследовательского отдела “Артериальной гипертензии”, м.н.с. Института трансляционной медицины, ORCID: 0000-0002-3664-5383, Звартау Н. Э. — к.м.н., руководитель отдела биомедицинских исследований лекарственных средств, с.н.с. научно-исследовательского отдела “Артериальной гипертензии”, начальник организационно-методического управления по кардиологии и ангиологии; с.н.с. Института трансляционной медицины, ORCID: 0000-0001-6533-5950, Дубинина Е. А. — к.психол.н., доцент кафедры клинической психологии и психологической помощи, н.с. лаборатории клинической психологии и психодиагностики, ORCID: 0000-0003-1740-6659, Хромов-Борисов Н. Н. — к.б.н., с.н.с., ORCID: 0000-0001-6435-7218, Конради А. О. — д.м.н., член-корреспондент Российской академии наук, зам. генерального директора по научной работе, директор Института трансляционной медицины, ORCID 0000-0001-8169-7812.

\*Автор, ответственный за переписку (Corresponding author): ionov\_mv@almazovcentre.ru

АГ — артериальная гипертензия, КЖ — качество жизни, связанное со здоровьем, КТТ — классическая теория тестов, ПИСП — показатели исходов, сообщаемых пациентами, СТТ — современная теория тестирования.

Рукопись получена 14.04.2019  
Рецензия получена 17.05.2019  
Принята к публикации 24.05.2019



Российский кардиологический журнал. 2019;24(7):40–46  
<http://dx.doi.org/10.15829/1560-4071-2019-7-40-46>

**Ключевые слова:** артериальная гипертензия, исходы, сообщаемые пациентами, качество жизни, связанное со здоровьем, опросник, классическая теория тестов, теория тестовых заданий.

## Hypertension specific patient-reported outcome measure. Part II: validation survey and item selection process

Ionov M. V.<sup>1,2</sup>, Zvartau N. E.<sup>1,2</sup>, Dubinina E. A.<sup>3,4</sup>, Khromov-Borisov N. N.<sup>1</sup>, Konradi A. O.<sup>1,2</sup>

**Aim.** Improvement of the health-related quality of life (HRQoL) is one of the basic principles of value-based medicine. HRQoL could be assessed by the patient-reported outcome measures (PROMs) also in case of arterial hypertension (HTN). However for HTN patients only generic PROMs are still used. Previously the group of experts had created the primary version of HTN-specific PROM. The purpose of the second part was to conduct a validation survey and to select the items in a statistically-based manner.

**Material and methods.** Validation survey was conducted in a large multidisciplinary center among patients with HTN stages 1-3 and healthy volunteers. Inclusion criteria were age >18 years old, ability to understand or complete the scale themselves, absence of significant illness requiring hospitalization. The items were selected according to the principles of classical test theory (CTT) and item response theory

(IRT). The criteria for CTT were sensitivity (standard deviation and coefficient of variation with corresponding confidence intervals), representativeness (item-total Pearson's correlation coefficient), internal consistency (Cronbach's  $\alpha$  coefficient). In IRT analysis two methods were adopted — value of four degrees of difficulty and the discrimination estimate. Each question was evaluated according to 8 criteria. An item was considered for selection when it was retained by  $\geq 4$  criteria. The expert panel considered practical significance of each item.

**Results.** A total of 430 questionnaires were distributed and 407 (94,7%) of them were returned completed (from 359 hypertensive patients, mean age  $62,3 \pm 1,7$  y.o.; 48 healthy volunteers, mean age  $38,8 \pm 10,5$  y.o.). The average time for PROM filling was  $24 \pm 4,2$  minutes. Of 163 questions, 27 met all 8 criteria and 3 questions did not match any. Of the 36 HTN-specific questions, 11 matched  $\geq 5$  criteria and

in the generic part there were 87 questions (33 in the PHY domain, 35 for PSY, 8 for SOC, 11 for THER). The symmetric distribution of criteria was seen in 25 questions, of which 11 were evaluated by experts and then retained. For 40 questions, <4 eligibility criteria were recorded, of which 9 were retained after expert review. The PROM draft contained 80 questions (19 questions in the physiology domain, 22 in psychology, 6 in social, 13 in therapy, 20 items are HTN-specific).

**Conclusion.** The methods of CTT and IRT allowed to reduce the PROM volume without losing the semantic richness and the need to reorganize the conceptual structure. The next step is the validation of the scale.

**Russian Journal of Cardiology. 2019;24(7):40–46**

<http://dx.doi.org/10.15829/1560-4071-2019-7-40-46>

**Key words:** arterial hypertension, patient-reported outcome measures, health-related quality of life, questionnaire, classical test theory, item response theory.

#### Информация о предыдущей публикации:

Ионов М. В., Звартау Н. Э., Дубинина Е. А., Конради А. О.

Болезнь-специфичный опросник по исходам, сообщаемым пациентами с артериальной гипертензией. Часть I: создание и первичная оценка

Российский кардиологический журнал. 2019;24(6):54–60

<http://dx.doi.org/10.15829/1560-4071-2019-6-54-60>

Артериальная гипертензия (АГ) как ведущая причина преждевременной смертности и инвалидности [1], по прогнозам сохранит лидирующую позицию вплоть до 2040 года [2]. Несомненно, основной целью здравоохранения является спасение и продление жизни, однако столь же важным следует признать стремление к улучшению ее качества, то есть снижению выраженности симптоматики, облегчению психологического и социального бремени заболеваний. Как и большинство хронических патологий, АГ оказывает влияние на качество жизни, связанное со здоровьем (КЖ) в диапазоне от незначительного до весьма существенного [3]. Этот факт вызывает особенный интерес при условии более чем миллиарда пациентов с АГ по всему миру, эффективность лечения у большинства из которых недостаточная [4].

Последние научные данные указывают на снижение уровня КЖ у пациентов с неконтролируемой АГ; однако даже при несомненной продуктивности некоторые действенные антигипертензивные препараты и рациональные схемы терапии могут также негативно влиять на этот показатель [5]. “Измерить” симптомы, влияние заболевания на психологические, социальные сферы жизни больного стало возможным с внедрением показателей исходов, сообщаемых пациентами (ПИСП) — высокоэффективным инструментом трансляции субъективных ощущений в объективную оценку.

Так как использование ПИСП в рутинной клинической практике позволяет изменить парадигму принятия решений, делая их более персонализированными, то ожидается, что медицинская помощь может быть качественнее и на популяционном уровне. Так, указания в международных и отечественных Рекомендательных документах по ведению пациентов с АГ носят общий характер и основаны на “грубой” стратификации пациентов по основным объектив-

**Conflicts of interest:** nothing to declare.

**Funding.** The study was conducted by a grant from the Russian Science Foundation (project № 17-15-01177).

<sup>1</sup>Almazov National Medical Research Center, St. Petersburg; <sup>2</sup>St. Petersburg National Research University of Information Technologies, Mechanics and Optics (ITMO University), St. Petersburg; <sup>3</sup>A. I. Herzen Russian State Pedagogical University, St. Petersburg; <sup>4</sup>V. M. Bekhterev National Medical Research Center of Psychiatry and Neurology, St. Petersburg, Russia.

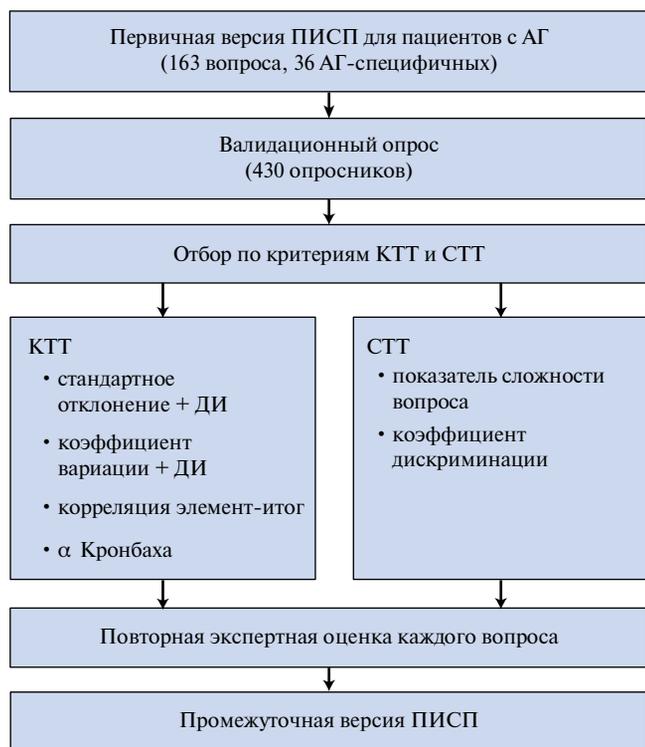
Ionov M. V. ORCID: 0000-0002-3664-5383, Zvartau N. E. ORCID: 0000-0001-6533-5950, Dubinina E. A. ORCID: 0000-0003-1740-6659, Khromov-Borisov N. N. ORCID: 0000-0001-6435-7218, Konradi A. O. ORCID 0000-0001-8169-7812.

**Received:** 14.04.2019 **Revision Received:** 17.05.2019 **Accepted:** 24.05.2019

ным показателям [6]. В свою очередь, ПИСП — более целостная и всеобъемлющая оценка КЖ и эффекта лечения с позиции пациента, а ПИСП-анализ может обеспечить врача ценной информацией для реализации принципов прецизионной медицины. Стоит также отметить, что порой результаты ПИСП-анализа являются первичными конечными точками клинических исследований, заменяя или дополняя “традиционные” объективные и лабораторные цели.

Подробные алгоритмы разработки инструментов для оценки КЖ даны в Рекомендательных документах международных агентств. Однако трудности выбора подходящего ПИСП отмечаются многими исследователями, так как зачастую он зависит от конкретной патологии, типа клинического испытания и постановки авторами/экспертами целей и задач. Как и с многими хроническими патологиями, пациенты с АГ пребывают в определенном континууме состояний, определяющихся степенью тяжести (смягчить которую можно заметно и в короткие сроки) и стадией (практически не поддающейся коррекции). Поэтому важно располагать надежными и краткими опросниками для пациентов с определенной патологией. Кроме того, оперируя адекватными данными, полученными с помощью болезнь-специфических ПИСП, становится возможным провести важный вариант экономического анализа — затраты-полезность (cost-utility analysis) [7]. Последний является одним из самых сложных и тонких методов экономического анализа, наиболее важен в концепции ценностно-ориентированного здравоохранения (клиническая, экономическая и пациент-ориентированная польза).

На первом этапе работы был описан процесс создания многомерного и многофакторного болезнь-специфического ПИСП для пациентов с АГ [8]. В рамках первой части исследования были проведены



**Рис. 1.** Дизайн второго этапа разработки ПИСП для пациентов с АГ. Массовый опрос и селекция вопросов.

**Сокращения:** ПИСП — показатель исходов, сообщаемых пациентами, АГ — артериальная гипертензия, КТТ — классическая теория тестирования, СТТ — современная теория тестирования, ДИ — доверительный интервал.

интервьюирование и пилотное анкетирование пациентов, последующая оценка состава опросника, что во многом сократило его объем. Нынешний этап работы нацелен на привлечение специальных статистических методов анализа психологических тестов, которые дополняют качественную экспертизу.

### Материал и методы

Исследование было выполнено в соответствии со стандартами надлежащей клинической практики (Good Clinical Practice) и принципами Хельсинкской Декларации. Протокол исследования был одобрен локальным Этическим комитетом. До включения в исследование у всех участников было получено письменное информированное согласие. В качестве методических рекомендаций по созданию и валидации ПИСП использовались руководства и документы Управления по санитарному надзору за качеством пищевых продуктов и медикаментов США (FDA) [9], Международного общества исследований в области фармакоэкономики и клинических исходов (ISPOR) [10], Международного общества по исследованиям качества жизни (ISOQOL) [11]. Исследование было выполнено при поддержке гранта Российского научного фонда (проект № 17-15- 01177).

**Валидационное исследование.** Опрос с первичной версией ПИСП [8] проводился в условиях амбулаторного отделения крупного многопрофильного медицинского центра.

В основную группу были последовательно включены пациенты с АГ 1-3 степеней тяжести, которые впервые оказались на приеме у специалиста-гипертензиолога (один из авторов статьи), при этом статус получаемой антигипертензивной терапии не учитывался. Вторую группу участников составили условно здоровые добровольцы.

Общими критериями включения были возраст не менее 18 лет, возможность понять цель и инструкции по заполнению, самостоятельно прочесть и ответить на вопросы печатной формы ПИСП.

Основными критериями невключения стали: оцениваемый субъективно врачом когнитивный дефицит или известный клинический диагноз энцефалопатии любой этиологии 2 и более высокой степеней тяжести, серьезная соматическая патология (сердечно-сосудистая или некардиальная), которая требовала стационарного (в том числе оперативного) лечения в ближайшее время. Всем участникам было предложено заполнить опросник после беседы с врачом и подписания формы информированного согласия.

**Статистический анализ.** Отбор вопросов проводился на основании классической теории тестов (classical test theory, КТТ) и современной теории тестирования (item response theory, СТТ).

При отборе вопросов на основании КТТ были использованы следующие 6 критериев их оценки:

1. Чувствительность (sensitivity) элемента определялась по стандартному отклонению. Вопрос мог рассматриваться как неподходящий, если значение стандартное отклонение (SD) было <1,0.

1.1. Отдельным критерием являлись показатели обеих границ 95% доверительного интервала стандартного отклонения и для приемлемости вопроса должны были иметь значения выше 1,0. Желательно, чтобы этому условию удовлетворяло не значение точечной оценки SD, а ее нижняя граница 95%.

2. Чувствительность вопроса оценивалась также посредством коэффициента вариации (CV) ответов; чтобы вопрос остался в промежуточной версии опросника, значение CV должно было превышать 20 единиц.

2.1. Подобно доверительному интервалу для стандартного отклонения, те же требования предъявлялись и к 95% доверительному интервалу коэффициента вариации, нижняя граница которого должна была находиться выше показателя 20 единиц.

3. Репрезентативность (representativeness) вопроса оценивалась на основании коэффициента корреляции “элемент-итог”. Вопрос считался подходящим, если показатель корреляции Пирсона между средним значением ответов вопроса и средним значением ответов его сферы превышал значение CV = 0,5.

4. Внутренняя согласованность (internal consistency) определялась исходя из показателя коэффициента  $\alpha$  Кронбаха, значение которого должно было превышать 0,5.

При отборе вопросов на основании принципов СТТ были использованы два критерия, значения которых оценивали методом максимума правдоподобия:

1. Коэффициент дискриминации  $a$  (discrimination estimate). Основной принцип: чем выше коэффициент  $a$ , тем выше информативность элемента. Если значение коэффициента превышало значение  $a > 0,5$ , то вопрос мог быть оставлен в опроснике.

2. степень сложности вопроса (value of degrees of difficulty), определяемая четырьмя коэффициентами  $b_1, b_2, b_3$  и  $b_4$  и удовлетворяющими неравенству  $b_1 < b_2 < b_3 < b_4$ . При этом значения крайних элементов  $b_1$  и  $b_4$  должны были находиться в диапазоне от  $-3$  до  $+3$ . В противном случае, вопросы, у которых значения этих элементов выпадали за пределы указанного диапазона, становились претендентами на удаление из опросника, поскольку распределение ответов на них будет смещено к одной из крайностей (ответы с баллами 1 или 5).

Каждый вопрос оценивался на основании восьми описанных выше критериев. Если элемент отвечал четырем и более из них, то он мог быть оставлен в структуре опросника. Кроме того, определялась важность и смысловая насыщенность вопроса на основании повторной оценки экспертной группой. Таким образом, посредством совместного применения статистического анализа и экспертных оценок формировались логика и информативность второй промежуточной версии ПИСП (рис. 1).

Статистическая обработка результатов была выполнена с помощью некоммерческого пакета с открытым кодом программ R Statistics (ver. 3.1.0, The R Foundation for Statistical Computing, Vienna, Австрия) и пакета программ SPSS (ver. 23.0, IBM, Chicago, IL, США). Уровень статистической значимости для различий был задан как  $p < 0,05$ . Использованы также специализированные программы: IRTShiny Version 1.1 (<http://kylehamilton.net/shiny/IRTShiny/>); Classical Test Theory (Item Analysis) — <http://kylehamilton.net/shiny/CTTShiny/> Для анализа надежности (Reliability Analysis) использовали программу jamovi (<https://www.jamovi.org/>).

### Результаты

В анкетировании участвовало 430 человек. Возвращено 407 заполненных ПИСП: 359 от пациентов с АГ: средний возраст  $62,3 \pm 11,7$  лет, 56,8% женщин; АГ 1 степени — 139 пациентов, АГ 2 степени — 136 пациентов, АГ 3 степени — 84 пациента; 48 от условно здоровых добровольцев (средний возраст  $38,8 \pm 10,5$  лет, 70,8% женщин). Эффективность заполнения опросника составила 94,7%, а среднее время запол-

нения составило  $24 \pm 4,2$  мин (для пациентов с АГ, так как здоровые добровольцы не отвечали на вопросы, касающиеся лечения). Все респонденты, которые заполнили опросник и с которыми удалось провести послетестовую беседу ( $n=128$ ), сообщили врачу-исследователю, что вопросы и варианты ответов были сформулированы корректно и не вызывали трудностей при анкетировании.

Анализ частотного распределения показал, что при заполнении ПИСП было 11,4% пропущенных ответов. Недостающие данные были проанализированы тестом Литтла на полностью случайные пропуски (Little's Missing Completely at Random Test):  $\chi^2=347$ ,  $p=0,39$ . Результаты теста показали, что распределение согласуется с нормальным, а пропуски являются случайными. Пропущенные данные были таким образом восстановлены методом множественного восстановления (multiple imputations).

Сформированная на основании понятийной структуры первичная версия опросника, состояла из общей (“неспецифической”) части, включавшей сферы “физиология” (PHY) с 43 вопросами (5 подсфер: физические симптомы, общее самочувствие и витальность, самооценка здоровья, ограничивающее влияние состояния физического здоровья, динамика физического здоровья), “психология” (PSY) — 42 вопроса (5 подсфер: эмоционально-поведенческие симптомы, когнитивные симптомы, общее психическое благополучие, ограничивающее влияние состояния психического здоровья, динамика психического здоровья), “социальной” сферы (SOC), которая содержала 15 вопросов (4 подсферы: социальная фрустрированность, социальные ресурсы, влияние физического и психического здоровья на социальную активность), и домена “лечения” (THER) с 27 элементами (6 подсфер: общая удовлетворенность лечением, физические изменения в связи с лечением, психологические изменения в связи с лечением, влияние режима лечения на повседневную жизнь, приверженность рекомендациям относительно режима лечения и образа жизни), и АГ-специфичной части, (13 вопросов в схожих с общей частью субдоменах PHY и THER, 4 вопроса — в PSY и 6 — в SOC).

Таким образом, всего было оценено 163 вопроса (из них 36 вопросов, касающихся исключительно АГ). Для каждого вопроса показатели критериев КТТ и СТТ представлены в таблице 1 (<http://dx.doi.org/10.15829/1560-4071-2019-7-116-125>, см. в конце эл. версии журнала на стр. 116-125). Из них соответствовали всем 8 критериям 27 вопросов (2 вопроса АГ-специфичных, 20 в сфере PHY, 4 в сфере PSY, 1 в сфере SOC). Несмотря на удовлетворительные показатели, вопросы PHY\_4\_3, PHY\_4\_5 и PHY\_4\_8 были удалены в связи с низкой практической значимостью, вопросы PHY\_4\_10 и PHY\_4\_12 исключены,

так как оказались дублирующими, а PHU\_4\_14 и PHU\_4\_17 не включались в промежуточную версию, потому как схожие вопросы имелись в АГ-части (HTN\_SOC\_5, HTN\_SOC\_7, соответственно). Ни одному из представленных критериев не отвечали 3 вопроса (THER\_6\_7 “Как часто Вы принимаете лекарства по совету знакомых или по собственному усмотрению без назначения врача?”, THER\_7\_1 “Как часто Вы пропускаете назначенные визиты к врачу?”, HTN\_THER\_14 “Как часто Вы питаетесь в заведениях быстрого питания (“фаст-фуд?”), и они были исключены из промежуточной версии опросника.

Из 36 вопросов АГ-части, 11 соответствовали  $\geq 5$  критериям; в общей части таких элементов оказалось 87 (из них 33 в домене PHU, 35 — в сфере PSY, 8 — в сфере SOC, 11 — в домене THER). Последовательная экспертная оценка позволила удалить дополнительно 3 вопроса из АГ-части (HTN\_SOC\_4, HTN\_THER\_10, HTN\_THER\_11), исходя из их недостаточной практической значимости. Из общей части были, дополнительно к уже упомянутым, исключены 28 элементов, из них: 7 PHU-вопросов (практическая незначимость, дублирование вопросов из АГ-части, расхождение с концептом АГ), 15 PSY-вопросов (низкая надежность, практическая незначимость), 3 SOC-вопроса (практическая незначимость, дублирование АГ-части), 3 THER-вопроса (практическая незначимость, дублирование АГ-части).

Симметричное распределение критериев наблюдалось у 25 вопросов (из них 9 АГ-специфичных); эти вопросы вновь оценивались авторами, в результате чего 11 (АГ — 5, PSY — 1, THER — 5 вопросов) были оставлены в промежуточной версии (в силу смыслов насыщенности, практической важности и для сохранения структуры концептуальной модели).

Для 40 вопросов было зарегистрировано менее 4 критериев (кроме трех вопросов, описанных выше) соответствия, 9 из них оставлены после экспертной оценки (7 вопросов АГ-части, вопросы PSY\_5\_5 и SOC\_1\_1 в соответствующих сферах, все вопросы исходя из их практической значимости).

Обоим основным статистическим критериям КТТ и СТТ (показатель надежности и трудность вопроса, соответственно) не соответствовали 46 вопросов (из них 21 вопрос в части АГ). Из этих элементов, для которых оба критерия не были выполнены, только 16 были оставлены в промежуточной версии шкалы (11 вопросов АГ-специфичных, по 2 вопроса в сферах PSY и THER, 1 вопрос в сфере SOC) благодаря их практической значимости и смыслов насыщенности.

С учетом значимого сокращения пула вопросов и для облегчения последующей процедуры валидации, АГ-специфичные вопросы были интегрированы в соответствующие сферы основной части ПИСП. Таким образом вопросы HTN\_PHU\_1-12 остались в субсфере “физические симптомы”; элементы HTN\_

PSY\_1-4 составили дополнительную субсферу “гипо- и гипернозогнозия” в домене PSY; вопрос HTN\_SOC\_3 дополнил субсферу “социальные ресурсы в лечении АГ”, а вопросы HTN\_SOC\_5,7 — субсферу “влияние физического здоровья на социальную активность”; вопросы HTN\_THER\_1,2 вошли в субдомен “общая удовлетворенность лечением”, HTN\_THER\_3-5 — в субсферу “приверженность лекарственной терапии”. Из промежуточной версии ПИСП были удалены субсферы “динамика физического здоровья”, “влияние психического здоровья на социальную активность”. В итоге, промежуточный вариант ПИСП обозначен 80 вопросами (19 вопросов в сфере PHU, 22 вопроса в сфере PSY, 6 — в сфере SOC, 13 — в сфере лечения (THER), а также 20 АГ-специфичных элементов) (Приложение 1, <http://dx.doi.org/10.15829/1560-4071-2019-7-126-129>, см. в конце эл. версии журнала на стр. 126-129).

### Обсуждение

Экспертный выбор, создание понятийной структуры и разработка будущего опросника являются одними из самых важных и сложных этапов. Однако, когда этот этап пройден, возникает необходимость отбора смыслов насыщенности, практически значимых, наиболее надежных вопросов. Важным этапом второй части исследования стала необходимость получить достаточный массив данных для статистического анализа.

Самым распространенным методом социальных и психологических исследований является массовый опрос. Это тем более важно при разработке новых или адаптации известных иноязычных ПИСП, так как вклад пациентов и следующее за заполнением интервьюирование — один из способов подтверждения содержательной валидности (content validity) [9]. Поэтому была предпринята попытка широкого охвата популяции пациентов с АГ, в особенности на амбулаторном этапе лечения в максимально приближенных к реальным условиям клинической практики. Тем не менее, массовость анкетирования неизбежно приводит к неполноте полученных данных, что зачастую связано с недостижимостью некоторых респондентов, усталостью или невнимательностью (при заполнении опросников большого объема), культурно-этническими и социальными особенностями индивидуумов в выборке [12]. В связи с большим объемом первичной версии ПИСП, вероятно, усталость или невнимательность респондентов явились основной причиной потери десятой доли требуемых данных, о чем свидетельствует результат теста MCAR Литтла.

Полученный банк данных ответов стал основой для оценки каждой единицы АГ-специфичного ПИСП по двум принципам психометрии — КТТ и СТТ, методы которых были использованы совместно.

Методы КТТ, которая также называется истинной теорией оценки (true score theory), легко доступны для понимания и использования. Основными критериями КТТ в данной работе считались стандартное отклонение и надежность (свобода от случайных ошибок). Стоит отдельно отметить, что авторы не оценивали факторную нагрузку вопроса. Это связано с значительным объемом вопросов, количеством сфер и субсфер, то есть вероятность получить недостоверный результат распределения по факторам была довольно высока. Для опросника Нурег-PRO эксплораторный факторный анализ был проведен на этапе первичного отбора вопросов [13], что, принимая во внимание небольшой изначальный кандидатный пул, было вполне резонно.

Следует отметить, что несмотря на признанность КТТ, в ней не учитываются латентные характеристики (latent trait) и способности/умения/навыки (ability) респондентов, и поэтому оценка надежности вопросов может быть неадекватной. В дополнение к КТТ, таким образом, была реализована оценка, базирующаяся на принципах СТТ — различия в кривых характеристик вопроса (item characteristics curves) и трудности вопросов. СТТ, являясь вероятностной и математически-обоснованной моделью, обладает тремя неоспоримыми преимуществами над КТТ: оценка дееспособности респондента не зависит от конкретного вопроса, механизм оценки не зависит от тестируемой популяции, точность механизма оценки способности респондента также может быть определена. При использовании СТТ можно определить нелинейную зависимость между ответом респондента и потенциальным его качеством или же описать связь между ответом и лежащим в основе вопроса фактором (скрытой характеристикой). Однако принципы СТТ довольно трудны в понимании и поэтому применение ограничено, в основном, разработкой учебных тестов [14].

Одно из важных соображений в рамках применения СТТ для разработки ПИСП касается надежности, размера и содержания. В среде специалистов растет понимание того, что комбинации качественных вопросов могут способствовать разработке наиболее валидных и кратких ПИСП, которые бы снизили “время респондента”. Для этого предпринимаются попытки не оценивать ПИСП как целостный концепт, а определять надежность и важность отдельных вопросов, основываясь на особенностях ответа пациентов. Активное внедрение СТТ и компьютер-

ных программ адаптивного тестирования позволило создать систему PROMIS (Patient-Reported Outcomes Measurement Information System), подбирающую из объемной базы вопросов минимальное количество максимально информативных элементов исходя из клинических характеристик пациента [15].

Большинство вопросов были удалены на основании комбинации методов КТТ и СТТ, однако значимые для общей структуры элементы, несмотря на несоответствие критериям у некоторых из них, были оставлены, исходя из оценки экспертной группой. Например, большинство несоответствующих вопросов АГ-части были оставлены, а вопросы заключительного субдомена THER (в общей и АГ-части) были исключены, значимо сокращена субсфера приверженности рекомендациям по изменению образа жизни. Элементы этой сферы разрабатывались на основании клинического суждения авторов и теоретической проблематики лечения пациентов с АГ (выбор ожидаемых вариантов врачебных рекомендаций, столь же объективно отражающих общую приверженность, как и прием антигипертензивной терапии). Вероятно, культурные, социологические и возрастные особенности, наряду с некоторым смещением выборки (более длительное течение АГ, известное пациенту направление в специализированный медицинский центр) могло повлиять на недостаточное соответствие предъявляемым критериям отбора.

### Заключение

Разработка болезнь-специфичного опросника по исходам, сообщаемым пациентами с АГ, прошла второй этап с применением методов классической и современной теорий тестирования, экспертными оценками. Полученные результаты позволили сократить его объем более чем вдвое за счет исключения неподходящих — дублирующих, ненадежных, сложных для понимания элементов. Кроме того, в промежуточной версии ПИСП сохранилась прежняя структура и концептуальная модель. Следующим этапом планируется проведение анализа валидности, надежности и чувствительности.

**Финансирование.** Исследование выполнено за счет гранта Российского научного фонда (проект № 17-15-01177).

**Конфликт интересов:** все авторы заявляют об отсутствии потенциального конфликта интересов, требующего раскрытия в данной статье.

## Литература/References

1. GBD 2017 Risk Factor Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Lond Engl*. 2018;392:1923-94. doi:10.1016/S0140-6736(18)32225-6.
2. Foreman KJ, Marquez N, Dolgert A, et al. Forecasting life expectancy, years of life lost, and all-cause and cause-specific mortality for 250 causes of death: reference and alternative scenarios for 2016-40 for 195 countries and territories. *The Lancet*. 2018;392:2052-90. doi:10.1016/S0140-6736(18)31694-5.
3. Trevisol DJ, Moreira LB, Kerkhoff A, et al. Health-related quality of life and hypertension: a systematic review and meta-analysis of observational studies. *J Hypertens*. 2011;29:179-88. doi:10.1097/HJH.0b013e328340d76f.
4. Forouzanfar MH, Liu P, Roth GA, et al. Global Burden of Hypertension and Systolic Blood Pressure of at Least 110 to 115 mm Hg, 1990-2015. *JAMA*. 2017;317:165-82. doi:10.1001/jama.2016.19043.
5. Youssef RM, Moubarak II, Kamel MI. Factors affecting the quality of life of hypertensive patients. *East Mediterr Health J Rev Sante Mediterr Orient Al-Majallah Al-Sihhiyah Li-Sharq Al-Mutawassit*. 2005;11:109-18.
6. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Eur Heart J*. 2018;39:3021-104. doi:10.1093/eurheartj/ehy339.
7. Yagudina RI, Sorovikov IV. Methodology of Cost-Utility Analysis in Pharmacoeconomic Studies. *Farmakoeconomika. Modern Pharmacoeconomic and Pharmacoepidemiology*. 2012;5:9-12. (In Russ.) Ягудина Р. И., Сороковиков И. В. Методология проведения анализа "Затраты-полезность" при проведении фармакоэкономических исследований. *Фармакоэкономика Современная Фармакоэкономика и Фармакоэпидемиология*. 2012;5:9-12.
8. Ionov MV, Zvartau NE, Dubinina EA, et al. Hypertension specific patient-reported outcome measure. Part I: development and primary evaluation *Russ J Cardiol*. 2019;24(6):54-60. (In Russ.) Ионов М. В., Звартан Н. Э., Дубинина Е. А., и др. Болезнь-специфичный опросник по исходам, сообщаемым пациентами с артериальной гипертензией. Часть I: создание и первичная оценка. *Российский кардиологический журнал*. 2019;24(6):54-60. doi:10.15829/1560-4071-2019-6-54-60.
9. US Department of Health and Human Services (USDHHS). Guidance for industry. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims> (13 April 2019).
10. Eremenco S, Coons SJ, Paty J, et al. PRO data collection in clinical trials using mixed modes: report of the ISPOR PRO mixed modes good research practices task force. *Value Health J Int Soc Pharmacoeconomics Outcomes Res*. 2014;17:501-16. doi:10.1016/j.jval.2014.06.005.
11. Reeve BB, Wyrwich KW, Wu AW, et al. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2013;22:1889-905. doi:10.1007/s11136-012-0344-y.
12. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, et al. Missing data and multiple imputation in clinical epidemiological research. *Clin Epidemiol*. 2017;9:157-66. doi:10.2147/CLEP.S129785.
13. Zhi L, Qiaojun L, Yanbo Z. Development and validation of patient-reported outcomes scale for hypertension. *Int J Qual Health Care J Int Soc Qual Health Care*. 2015;27:369-76. doi:10.1093/intqhc/mzv060.
14. Bock RD. A Brief History of Item Theory Response. *Educ Meas Issues Pract* 1997;16:21-33. doi:10.1111/j.1745-3992.1997.tb00605.x.
15. Bjorner JB, Rose M, Gandek B, et al. Difference in method of administration did not significantly impact item response: an IRT-based analysis from the Patient-Reported Outcomes Measurement Information System (PROMIS) initiative. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2014;23:217-27. doi:10.1007/s11136-013-0451-4.