

**ПРАВИТЕЛЬСТВО МОСКВЫ
ДЕПАРТАМЕНТ ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ**

**ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ЗДРАВООХРАНЕНИЯ
ГОРОДА МОСКВЫ «НАУЧНО-ПРАКТИЧЕСКИЙ КЛИНИЧЕСКИЙ ЦЕНТР
ДИАГНОСТИКИ И ТЕЛЕМЕДИЦИНСКИХ ТЕХНОЛОГИЙ ДЕПАРТАМЕНТА
ЗДРАВООХРАНЕНИЯ ГОРОДА МОСКВЫ»**

СОГЛАСОВАНО

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


С.П. Морозов

«24» июня 2019 года

РЕКОМЕНДОВАНО

Экспертным советом по науке
Департамента здравоохранения
города Москвы № 8



«25» июня 2019 года

**КЛИНИЧЕСКИЕ ИСПЫТАНИЯ
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ
ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ
(ЛУЧЕВАЯ ДИАГНОСТИКА)**

Методические рекомендации № 43

ISSN 2618-7124

УДК 615.84+616-073.75

ББК 53.6

К-49

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

Составители:

Морозов С.П. – д.м.н., профессор, главный внештатный специалист по лучевой и инструментальной диагностике МЗ РФ в ЦФО РФ, главный внештатный специалист по лучевой и инструментальной диагностике ДЗМ, директор ГБУЗ «НПКЦ ДиТ ДЗМ»

Владимирский А.В. – д.м.н., заместитель директора по научной работе ГБУЗ «НПКЦ ДиТ ДЗМ»

Кляшторный В.Г. – к.б.н., аналитик отдела координации научной деятельности ГБУЗ «НПКЦ ДиТ ДЗМ»

Андрейченко А.Е. – к.ф.-м.н., старший научный сотрудник отдела разработки средств контроля и технического мониторинга ГБУЗ «НПКЦ ДиТ ДЗМ»

Кульберг Н.С. – к.ф.-м.н., руководитель отдела разработки средств медицинской визуализации ГБУЗ «НПКЦ ДиТ ДЗМ»

Гомбоевский В.А. – к.м.н., руководитель отдела развития качества радиологии ГБУЗ «НПКЦ ДиТ ДЗМ»

К-49 Морозов С.П., Владимирский А.В., Кляшторный В.Г., Андрейченко А.Е., Кульберг Н.С., Гомбоевский В.А. Клинические испытания программного обеспечения на основе интеллектуальных технологий (лучевая диагностика). Препринт № ЦДТ-2019-1 / Серия «Лучшие практики лучевой и инструментальной диагностики». – Вып. 23. – М., 2019. – 33 с.

Предназначение: обеспечить методологическую базу для проведения клинических испытаний алгоритмов и программного обеспечения, разработанных на основе интеллектуальных технологий («искусственного интеллекта»). Подобные испытания рассматриваются как этап подготовки к регистрации программного решения в качестве медицинского изделия. Рассматриваются подходы к оценке точности и эффективности разработок на основе «искусственного интеллекта» для лучевой диагностики.

Данный документ является собственностью Департамента здравоохранения города Москвы, не подлежит тиражированию и распространению без соответствующего разрешения. При использовании материала ссылка на источник обязательна

Для корреспонденции: info@npcmr.ru, npcmr@zdrav.mos.ru
109029, г. Москва, Средняя Калитниковская улица, дом 28, стр. 1
+7 (495) 276-04-36

© Департамент здравоохранения города Москвы, 2019

© ГБУЗ «НПКЦ ДиТ ДЗМ», 2019

© Коллектив авторов, 2019

СОДЕРЖАНИЕ

Нормативные ссылки.....	4
Определения.....	5
Обозначения и сокращения.....	7
Введение.....	8
Основная часть. Общие положения.....	10
Этапы клинической оценки.....	13
Метрики для клинической оценки.....	15
Требования к эталонным данным.....	20
Оформление результатов.....	21
Заключение.....	23
Список использованных источников.....	24
Приложение.....	25



НОРМАТИВНЫЕ ССЫЛКИ

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

1. Федеральный закон от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации».
2. Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных».
3. Федеральный закон от 27.07.2006 № 149-ФЗ «Об информации, информационных технологиях и о защите информации».
4. Федеральный закон от 27.12.2002 № 184-ФЗ «О техническом регулировании».
5. Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы».
6. Постановление Правительства Российской Федерации от 27.12.2012 № 1416 «Об утверждении Правил государственной регистрации медицинских изделий».
7. Постановление Правительства Российской Федерации от 05.05.2018 № 555 «О единой государственной информационной системе в сфере здравоохранения».
8. Приказ Министерства здравоохранения Российской Федерации от 09.01.2014 № 2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».
9. Приказ Министерства здравоохранения Российской Федерации от 01.04.2016. № 200н «Об утверждении правил надлежащей клинической практики».
10. Приказ Департамента здравоохранения города Москвы от 25.12.2017 № 918 «О регламенте регистрации данных в системе «Единый радиологический информационный сервис» в медицинских организациях государственной системы здравоохранения города Москвы».
11. ГОСТ Р ИСО 14155-2014 «Клинические исследования. Надлежащая клиническая практика» (приказ Ростехрегулирования от 04.06.2014 № 497-ст).
12. Письмо Федеральной службы по надзору в сфере здравоохранения от 30.12.2015 № 01И-2358/15 «О регистрации программного обеспечения».
13. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы / М.: ФГБУ «ВНИИИМТ» Росздравнадзора, 2018. – 31 с.

ОПРЕДЕЛЕНИЯ

В настоящем документе применены следующие термины с соответствующими определениями:

База медицинских данных (эталонные данные) – это упорядоченная совокупность медицинских данных, включая диагностические изображения и сведения о наличии патологических изменений на изображениях; структурированные случаи оказания медицинской помощи и связанные с ними электронные медицинские документы из электронных медицинских карт; библиотеки ключевых слов, словосочетаний и их критичных сочетаний, данные генетических тестов, а также включая комбинации различных медицинских данных, объединенные вокруг одного деперсонифицированного пациента (т.н. «цифровой двойник» – «digital twin»). Если база содержит подтвержденные сведения об окончательном диагнозе для каждого случая или подтвержденном патологическом процессе в изображениях, то она называется «верифицированной».

Интеллектуальные технологии – информационные технологии, созданные на основе «искусственного интеллекта».

«Искусственный интеллект» (ИИ) – 1. Научное направление, в рамках которого изучаются проблемы аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными; входит в комплекс компьютерных наук; создаваемые технологии относятся к информационным технологиям. 2. Комплекс технологических и программных решений, приводящих к результату, аналогичному интеллектуальной деятельности человека, и используемых для решения прикладных задач с помощью систем компьютерного зрения, обработки естественного языка, распознавания и синтеза речи, рекомендательных систем и интеллектуальных систем поддержки принятия решений, а также систем, основанных на перспективных методах.

Математическая модель – это абстракция реального явления, в которой интересующие исследователя отношения между реальными элементами заменены подходящими, зачастую упрощенными, отношениями между математическими объектами. Иными словами, это представление реальной ситуации с помощью математического языка.

Машинное обучение – 1. Раздел компьютерных наук, изучающий применение математических и статистических методов для создания обучающихся информационных систем. Такие системы обучаются решать основную задачу путем



предварительного решения множества сходных задач. Является частью «искусственного интеллекта». 2. Технологии автоматического обучения алгоритмов ИИ распознаванию и классификации на тестовых выборках объектов для повышения качества распознавания, обработки и анализа данных, прогнозирования.

Порог активации («cut-off», точка отсечения) – значение оцениваемого параметра, при достижении которого принимается решение о наличии и/или отсутствии патологии.

Разметка (тегирование) – процесс внесения в базу данных стандартизированных сведений о наличии патологических изменений на диагностических изображениях, сведений о локализации и характере патологии, информации о верификации, извлеченных из неструктурированных текстовых записей признаков, иных связанных данных; выполняется для каждого изображения (исследования) или структурированного электронного медицинского документа независимо несколькими квалифицированными врачами-специалистами.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

ЕРИС – Единый радиологический информационный сервис

ИИ – «искусственный интеллект»

ИТ – информационные технологии

МО – медицинская организация

ПО – программное обеспечение

РИС – радиологическая информационная система

РФ – Российская Федерация

ЦАМИ – централизованный архив медицинских изображений

AUC – Area Under Curve (площадь под характеристической кривой)

CE – Conformité Européenne (маркировка «европейское соответствие»)

CONSORT – Consolidated Standards of Reporting Trials (диаграмма потока участников исследования)

FDA – Food and Drug Administration (Управление по санитарному надзору за качеством пищевых продуктов и медикаментов)

PACS – Picture Archiving and Communication System (система хранения и обмена изображениями)

ROC – Receiver Operating Characteristic (характеристическая кривая)

FROC – Free Response Operating Characteristic (характеристическая кривая со свободным ответом)



ВВЕДЕНИЕ

«Искусственный интеллект» – это научное направление, в рамках которого изучаются проблемы аппаратного или программного моделирования тех видов человеческой деятельности, которые традиционно считаются интеллектуальными; входит в комплекс компьютерных наук; создаваемые технологии относятся к информационным технологиям. Научной новизны в концепции «искусственного интеллекта» мало. Автоматизированный анализ медицинской информации – это давно существующий раздел науки и практики. Однако в последние годы наблюдается взрывной рост нового поколения интеллектуальных технологий, он связан со значительным инженерным прогрессом вычислительных возможностей компьютерной техники и математики. «Искусственный интеллект» применим для решения ряда задач в медицинской диагностике (поддержка принятия решений, морфометрия, автоматизация участков производственных процессов, контроль качества и т.д.). Для практического применения системы, созданные на основе интеллектуальных технологий, должны пройти клинические испытания, оценку диагностической точности и получить официальный статус медицинского изделия. В соответствии с Федеральным законом от 21.11.2011 №323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» медицинскими изделиями являются любые инструменты, аппараты, приборы, оборудование, материалы и прочие изделия, применяемые в медицинских целях отдельно или в сочетании между собой, а также вместе с другими принадлежностями, необходимыми для применения указанных изделий по назначению, включая специальное программное обеспечение, и предназначенные производителем для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека, проведения медицинских исследований, восстановления, замещения, изменения анатомической структуры или физиологических функций организма, предотвращения или прерывания беременности, функциональное назначение которых не реализуется путем фармакологического, иммунологического, генетического или метаболического воздействия на организм человека.

Программное обеспечение (ПО), оказывающее влияние на принятие врачом клинически значимых решений, либо формирующее новую клинически значимую информацию, несет потенциальный риск применения и тем самым – причинения вреда здоровью пациента. В этой связи такое ПО, включая созданное с использованием технологий «искусственного интеллекта», подлежит регистрации в качестве медицинского изделия. Обязательным подготовительным этапом перед направлением такого ПО на регистрацию в качестве медицинского изделия является проведение испытаний клинической эффективности. Данные методические рекомендации устанавливают единообразные подходы к опреде-



лению диагностической эффективности такого ПО в процессе клинических испытаний. Причем, программное обеспечение на основе ИИ может применяться как самостоятельное медицинское изделие, так и быть принадлежностью (составной частью) другого медицинского изделия или предназначаться для установки (инсталляции) на этом медицинском изделии.

ОСНОВНАЯ ЧАСТЬ

Общие положения

«Искусственный интеллект» в диагностике представляет мультидисциплинарное (медицина, биология, математика и компьютерные науки) прикладное научное направление по постановке и решению задач моделирования отдельных компонентов аналитических процессов, являющихся частью интеллектуальной деятельности врача (табл.1).

Таблица 1. Типовые формы применения «искусственного интеллекта» в лучевой диагностике

Основные задачи (вопросы)	Действие ИИ	Типовая медицинская задача
Объект присутствует или нет	Обнаружение (детекция)	Скрининг, массовые профилактические осмотры
Определение типа объекта	Классификация	Определение характера патологии
Определение размеров, характеристика объекта	Сегментация, определение радиомических особенностей	Медицинская морфометрия (мониторинг течения патологического процесса, автоматическое формирование описания диагностического изображения)
Распознавание речи и текста человека	Обработка естественного языка	Протоколирование, контроль качества

Любой алгоритм «искусственного интеллекта», успешно прошедший клинические испытания, функционирует исключительно под контролем медицинских работников в качестве системы поддержки принятия врачебных решений. Генерируемые алгоритмом результаты могут использоваться или не использоваться медицинскими работниками. Применение «искусственного интеллекта» направлено на повышение эффективности системы здравоохранения и производительности труда персонала, снижение рисков и ошибок, стандартизацию представления результатов диагностических исследований.

Программное обеспечение, разработанное на основе интеллектуальных технологий, проходит клиническую апробацию на этапе подготовки к регистрации в качестве медицинского изделия (в том числе в соответствии со статьей 36.1. Федерального закона от 21.11.2011 № 323-ФЗ). В процессе испытаний должна быть установлена диагностическая эффективность, подтверждены безопасность использования и соответствие реальных характеристик ПО тем, которые были заявлены изготовителем.

В соответствии со статьей 38 Федерального закона от 21.11.2011 № 323-ФЗ «Об основах охраны здоровья граждан в Российской Федерации» медицинское изделие – это «любые инструменты, аппараты, приборы, оборудование, материалы и прочие изделия, применяемые в медицинских целях отдельно или в сочетании между собой, а также вместе с другими принадлежностями, необходимыми для применения указанных изделий по назначению, включая специальное программное обеспечение, и предназначенные производителем для профилактики, диагностики, лечения и медицинской реабилитации заболеваний, мониторинга состояния организма человека, проведения медицинских исследований, восстановления, замещения, изменения анатомической структуры или физиологических функций организма, предотвращения или прерывания беременности, функциональное назначение которых не реализуется путем фармакологического, иммунологического, генетического или метаболического воздействия на организм человека. Медицинские изделия могут признаваться взаимозаменяемыми, если они сравнимы по функциональному назначению, качественным и техническим характеристикам и способны заменить друг друга».

В соответствии с информационным письмом Федеральной службы по надзору в сфере здравоохранения, медицинским изделием, подлежащим государственной регистрации в установленном порядке, является программное обеспечение для :

- управления работой оборудования и мониторинга за его функционированием;
- получения от оборудования диагностических данных, их накопления и расчета в автоматическом режиме;
- мониторинга функций организма человека и передачи полученных данных (в том числе посредством беспроводных технологий);
- расчета параметров подбора дозы (облучения, лекарственного средства, рентгеноконтрастного вещества и т.д.);
- обработки данных, полученных с диагностического медицинского оборудования, передачи их на системы планирования и терапии;
- обработки медицинских изображений (включая изменение его качества, цветового разрешения и т.д.);
- 3D-моделирования (в том числе, в стоматологии и зуботехнике, протезировании, ортопедии, имплантологии, макетировании органов и костей и т.п. [2]);
- связи диагностического и лечебного оборудования;
- обработки цифровых изображений (в том числе с получением данных от диагностического оборудования в неизменном виде).

Дополнительно следует учитывать, что ПО относится к медицинским изделиям в случае, если оно предназначено для [2]:



- обработки, хранения и передачи медицинской документации;
- оказания медицинской помощи с применением телемедицинских технологий;
- установки в специальной информационной системе оказания медицинской помощи на дому;
- обработки данных для электронной регистрации медицинского оборудования, что позволяет проводить учет применяемых в больнице медицинских изделий;
- создания канала связи между пациентом и медицинским специалистом, чтобы пациент мог проводить самостоятельное лечение в соответствии с правилами, определенными в процессе использования приложения, и/или для обеспечения мониторинга населения;
- установки в информационной системе банка крови, или существующих электронных вычислительных машинах, или децентрализованных компьютеров/сетях.

Предварительная опытная эксплуатация

До этапа получения программным обеспечением статуса медицинского изделия допустимо применение «искусственного интеллекта» в рамках пилотных проектов, научно-исследовательских работ и апробаций. К подобной деятельности следует допускать только компании, алгоритмы которых соответствуют следующим **критериям** (список и анкета – см. Приложение 1). Критерии вносятся в техническое задание, описывающее требования к ПО на основе интеллектуальных технологий. Также в техническом задании обязательно должны быть указаны **цели применения** «искусственного интеллекта» (например, на основе табл.1) по отдельным модальностям, производственным процессам.

Таким образом, применяемые в лучевой диагностике алгоритмы и программное обеспечение на основе «искусственного интеллекта» относятся к медицинским изделиям и должны проходить клинические испытания. При этом программное решение может применяться самостоятельно или совместно с другими медицинскими изделиями (как аппаратурой, так и программным обеспечением, например, ЕРИС, PACS, РИС и т.д.). Предварительная опытная эксплуатация ИИ возможна при условии соответствия компании-разработчика указанным выше методическим критериям.

¹ Письмо Федеральной службы по надзору в сфере здравоохранения от 30.12.2015 № 01И-2358/15 «О регистрации программного обеспечения»

Этапы клинической оценки

Процесс клинических испытаний организуют в соответствии с действующей нормативно-правовой базой и рекомендованной методикой экспертизы качества, эффективности и безопасности медицинских изделий в части программного обеспечения. В ходе испытаний ПО на основе интеллектуальных технологий должна быть дана «оценка клинической связи (научной обоснованности) степени адекватности подбора данных и системы их анализа (концепция, измерения, заключение) целевому назначению» [2].

Клиническая оценка ПО, разработанного именно на основе «искусственного интеллекта» и предназначенного для применения в лучевой диагностике, осуществляется в 2 этапа (рис.1):

1. Аналитическая валидация – оценка эффективности путем использования эталонных данных.
2. Клиническая валидация – оценка эффективности путем использования в рамках стандартного производственного процесса.

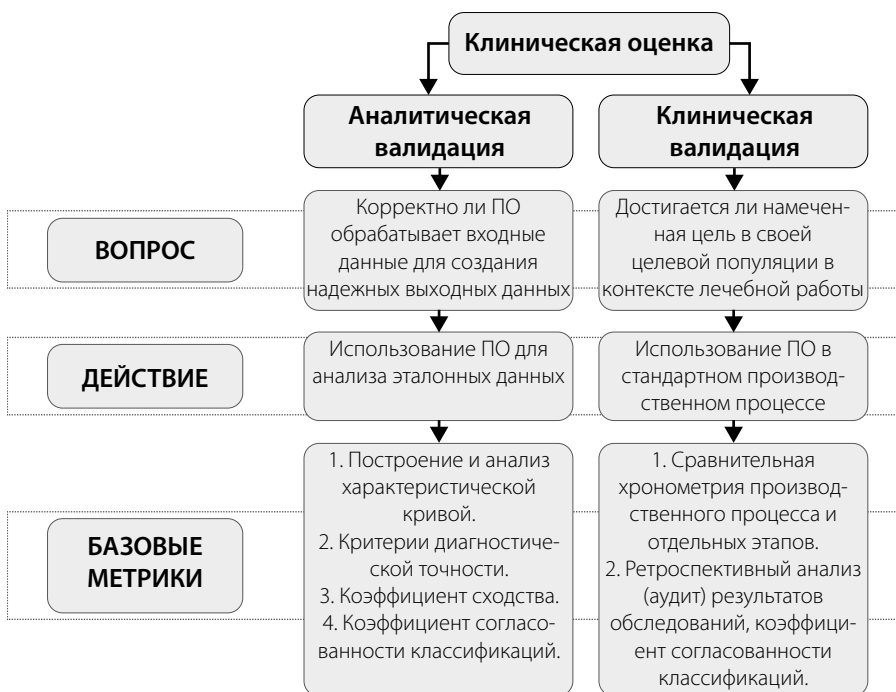


Рисунок 1 – Схема проведения клинической оценки программного обеспечения на основе интеллектуальных технологий (для сферы лучевой диагностики)



На первом этапе определение клинического значения осуществляется путем генерации и анализа новых диагностических данных для конкретного предполагаемого использования ПО:

- формируется база медицинских данных;
- осуществляется автоматизированный анализ базы (как имитация диагностического процесса);
- результаты анализа сопоставляются с эталоном – разметкой базы медицинских данных;
- проводится математический и статистический анализ.

Значимость первого этапа в клинической оценке ПО на основе ИИ составляет 90-95%.

На втором этапе ПО применяется по целевому назначению в рамках типового производственного процесса. Процедуры, выполняемые с применением ПО на основе «искусственного интеллекта», оцениваются с позиций производительности и качества. Соответственно, проводится сопоставительный хронометраж (при этом измерения должны проводиться многократно, с участием различных сотрудников-исполнителей, в разное время суток и периоды рабочих смен) и ретроспективный пересмотр результатов обследований. Организаторами испытания должны быть предусмотрены меры по минимизации рисков и обеспечению безопасности (безвредности для пациентов и их законных представителей) производственных процессов. Значимость второго этапа в клинической оценке ПО на основе ИИ составляет 5-10%.

В программе проведения клинических испытаний должны быть указаны:

- детальные сведения о базе медицинских данных (эталонных данных), включая номер свидетельства о государственной регистрации, количество и характер данных, сведения о разметке (см. далее);
- сведения о необходимости инсталляции ПО с применением технических средств испытательной медицинской организации;
- сведения об использовании вычислительных мощностей, предоставленных разработчиком ПО;
- сведения о деинсталляции ПО после завершения испытаний (в связи с тем, что ПО устанавливается и применяется на соответствующем оборудовании до получения регистрационного удостоверения).

В целом, формирование программы рекомендуется осуществлять в соответствии с чек-листом (см. приложение 2).

Метрики для клинической оценки

Аналитическая валидация

Аналитический этап осуществляется в дизайне диагностического исследования (см. приложение), при этом используется стандартный набор метрик.

Автоматизированный анализ медицинских данных посредством исследуемого ПО на основе ИИ считают индекс-тестом.

Результаты разметки базы медицинских данных в соответствии с известной методологией – референс-тестом.

Для сопоставления результатов индекс- и референс-теста составляют четырехпольную таблицу и определяют абсолютное значение для каждого вида результатов (табл.2). Далее производят выбор и расчет релевантных метрик (табл.3) с определением достоверности в рамках 95% доверительного интервала.

Таблица 2. Виды результатов применения ИИ в соответствии с четырехпольной таблицей

Вид результата			
Истинно-положительный	Истинно-отрицательный	Ложно-положительный	Ложно-отрицательный
ИП	ИО	ЛП	ЛО
ПО выявило патологию при ее наличии	ПО не выявило патологию при ее отсутствии	ПО выявило патологию при ее отсутствии	ПО не выявило патологию при ее наличии

Таблица 3. Типовой выбор метрик для аналитического этапа исследования

Действие ИИ	Базовые метрики
Обнаружение (детекция)	Стандартный набор метрик
Классификация	Стандартный набор метрик
Сегментация	Коэффициент сходства
Обработка естественного языка	Коэффициент согласованности классификаций

В зависимости от решаемой задачи могут применяться различные сочетания метрик. Например, анализ характеристической кривой и определение коэффициента согласованности классификаций и т.д. Такие основные показатели, как чувствительность, специфичность, площадь под характеристической кривой должны быть использованы для характеристики каждого типового действия «искусственного интеллекта».

Стандартный набор метрик

Для сопоставления эффективности индекс-теста относительно референс-теста применяется классический набор показателей, характеризующих диагностическую результативность (табл.4).

Таблица 4. Стандартный набор метрик диагностической ценности

Показатель	Значение	Формула
Чувствительность	Вероятность того, что индекс-тест окажется позитивным при наличии признака	$ИП / (ИП + ЛО)$
Специфичность	Вероятность того, что индекс-тест окажется негативным при отсутствии признака	$ИО / (ИО + ЛП)$
Точность (общая валидность)	Уровень соответствия полученных результатов индекс-тест к действительным значениям референс-теста	$(ИП + ИО) / (ИП + ИО + ЛП + ЛО)$
Отношение правдоподобия положительного результата	Математически отображает во сколько раз при положительном результате индекс-теста вероятность наличия целевой патологии превышает вероятность его отсутствия	$чувствительность / (1 - специфичность)$
Отношение правдоподобия отрицательного результата	Математически отображает во сколько раз при отрицательном результате индекс-теста вероятность наличия целевой патологии превышает вероятность его отсутствия	$(1 - чувствительность) / специфичность$
Прогностическая ценность положительного результата	Вероятность заболевания при положительном результате индекс-теста	$ИП / (ИП + ЛП)$
Прогностическая ценность отрицательного результата	Вероятность отсутствия заболевания при отрицательном (нормальном) результате индекс-теста	$ИО / (ИО + ЛО)$

Все метрики, кроме отношения правдоподобия, **оцениваются** в диапазоне 0-1 или в процентах от 0 до 100%:

Оценка
<0,6 – непригодно
0,61 - 0,8 – требуется доработка
>0,81 – может быть допущено к клинической валидации

Отношение правдоподобия положительного результата должно быть как можно выше, тогда как отношение правдоподобия отрицательного результата - как можно ниже.

Характеристическая кривая ((F)ROC-кривая)

Характеристическая кривая – отображение соотношения между долей объектов от общего количества носителей признака, верно классифицированных как несущих признак (то есть чувствительностью) и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущих признак (то есть специфичностью) при варьировании порога активации. Метрика диагностической ценности: площадь под кривой (AUC) - площадь, ограниченная (F)ROC-кривой и абсциссой.

Классическая ROC-кривая: график зависимости чувствительности от 1-специфичности (по оси абсцисс – 1-специфичность, по оси ординат – чувствительность).

В дополнение к классической рекомендуется построение ROC-кривой, основанной на прогностической ценности: график зависимости прогностической ценности положительного результата от 1- прогностическая ценность отрицательного результата (по оси абсцисс – 1- прогностическая ценность отрицательного результата, по оси ординат – прогностическая ценность положительного результата). Такая ROC-кривая может давать преимущества в некоторых случаях, поскольку учитывает распространенность изучаемого признака в популяции.

При анализе указанных ROC-кривых необходимо провести определение оптимальных значений порога активации (так называемого «cut-off»). Для этого рекомендуется использовать две методики:

- 1) минимальное расстояние от верхнего левого угла до ROC-кривой (Минимум d);
- 2) индекс Юдена (Youden index), который отображает максимальное расстояние от диагональной линии до ROC-кривой (рис.2). Приоритет той или иной методи-

ке определения «cut-off» должен отдаваться в зависимости от целей исследования.

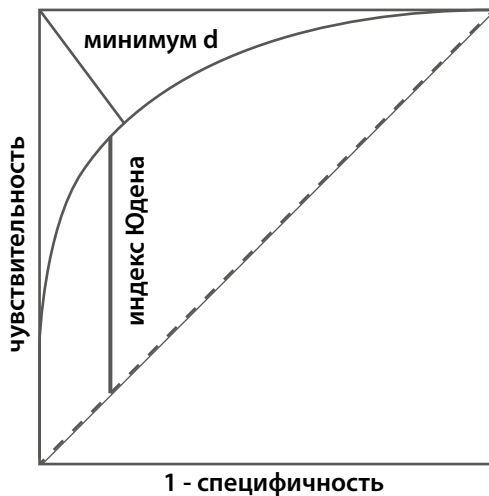


Рисунок 2 – Методика определения оптимальных значений порога активации «cut-off»

ROC-кривая со свободным ответом (FROC-кривая): график зависимости чувствительности от ЛП (по оси абсцисс – количество ложно-положительных случаев, по оси ординат – чувствительность).

Площадь под кривой (AUC) **оценивается** в диапазоне 0-1:

Оценка
<0,6 – непригодно
0,61 - 0,8 – требуется доработка
>0,81 – может быть допущено к клинической валидации

Коэффициент согласованности классификаций

Согласованность (конкордантность) – математическое отображение согласованности классификации двух экспертов относительно одного явления.

Формула расчета по методу Коэна (каппа Коэна) (Cohen J., 1960):

$$K = (P_o - P_e) / (1 - P_e),$$

где P_0 – доля случаев, когда измерения совпадали, и P_e – ожидаемая доля случаев случайного совпадения. Порядок расчета:

		Эксперт №1		
		1	2	Всего
Эксперт №2	1	P_{11}	P_{12}	$P_{1_}$
	2	P_{21}	P_{22}	$P_{2_}$
	Всего	$P_{_1}$	$P_{_2}$	1

$$P_0 = P_{11} + P_{22}$$

$$P_e = P_{_1}P_{1_} + P_{_2}P_{2_}$$

Оценивается в диапазоне 0-1 или в процентах от 0 до 100%:

Оценка
<0,6 – непригодно
0,61 - 0,8 – требуется доработка
>0,81 – может быть допущено к клинической валидации

Кoeffициент сходства

Бинарный показатель для количественного определения степени сходства биологических объектов. Формула расчета по методу Дайса-Сёренсена (Dice L., 1945, Sörensens T., 1948):

$$DSC = 2|A \cap B| / (|A| + |B|)$$

где в числителе – удвоенное количество совпавших оценок, в знаменателе – общее количество оценок.

Для множеств с одинаковой мощностью **оценивается** в диапазоне 0-1:

Оценка
<0,6 – непригодно
0,61 - 0,8 – требуется доработка
>0,81 – может быть допущено к клинической валидации

Клиническая валидация

Хронометраж – метод оценки результативности применения ПО, фиксация и замеры продолжительности выполняемых действий (процедур, операций) в рамках производственных процессов. Например, продолжительность описания цифровой рентгенографии легких, выполненной в профилактических целях с и без «искусственного интеллекта».

Аудит результатов обследований – ретроспективный анализ качества интерпретации медицинских данных, например, проведения и описания лучевых исследований, выполненных с применением интеллектуальных технологий, с целью определения степени соответствия рекомендуемым стандартам [4]. Аудит позволяет оценить качество оказания медицинской помощи, включая проведение и интерпретацию результатов диагностических исследований с и без «искусственного интеллекта», достоверность различий определяется статистически.

Требования к эталонным данным

Для выполнения аналитического этапа клинических испытаний применяют эталонные данные – базу медицинских данных.

База медицинских данных – это упорядоченная совокупность: диагностических изображений одной модальности и/или однотипных медицинских документов; сведений о наличии и характере патологических изменений на изображениях; для текстовых документов – библиотеки ключевых слов, словосочетаний и их критичных сочетаний. Если база содержит сведения об окончательном диагнозе для каждого случая, то она называется «верифицированной». Общеупотребительные синонимы: «выборка», «data-set», «тегированная база».

База медицинских данных применяется для аналитической валидации программного обеспечения, разработанного на основе ИИ для решения задач медицинской диагностики.

Эталонная база данных должна быть проверена профильной медицинской научно-исследовательской организацией на предмет полноты и качества содержащейся в ней информации.

Рекомендуется при проведении клинических испытаний применять эталонные данные, имеющие государственную регистрацию базы данных.

Эталонная база медицинских данных должна содержать такие сведения (описательного характера):

- 1) номер свидетельства о государственной регистрации базы данных (рекомендательно);
- 2) характеристика популяции (гендерно-возрастные показатели, этнический состав, регионы проживания и т.д.); сведения о деперсонализации; сведения о

медицинских организациях, послуживших источниками для формирования базы данных;

3) характеристика исследований: анатомическая область(и), модальность, проекции, типы медицинских изделий - диагностических приборов, виды и характеристики протоколов исследований;

4) целевая патология в соответствии с Международной классификацией болезней;

5) общее количество клинических случаев, исследований, изображений, документов и их распределение по диагностическим группам;

6) соотношение случаев «норма»:«патология» (случаи «патология» могут быть разделены на несколько подклассов);

7) сведения о верификации (патогистологическом или ином окончательном диагнозе);

8) методология разметки (ранее представленной в виде научных публикаций, методических рекомендаций или патентов).

Эталонная база медицинских данных должна отвечать таким требованиям:

1) соотношение «норма»:«патология» в эталонной базе должно соответствовать распространенности целевой патологии в популяции;

2) при формировании эталонной базы должны использоваться данные из нескольких медицинских организаций;

3) демографические, социально-экономические характеристики и основные показатели здоровья пациентов, чьи данные включаются в эталонную базу данных, должны соответствовать усредненным характеристикам популяции территории, на которой планируется использование ИИ;

4) планируемый размер эталонной базы данных должен быть обоснован в протоколе исследования, исходя из статистических соображений и желаемой точности оценки основных метрик, указанных выше.

Оформление результатов

При оценке результатов клинических испытаний следует учитывать следующее:

- ПО проверяется с применением биоматериалов и всего необходимого оборудования (в том числе, других медицинских изделий), предназначенного для обеспечения надлежащей работы исследуемого ПО;
- в процессе проведения клинических испытаний должны быть задействованы все модули ПО (при модульной архитектуре), опробованы все функции, указанные производителем в эксплуатационной документации;
- должны быть оценены все позиции доказательных баз ПО (клинические данные, анализ взаимозаменяемых медицинских изделий и т.п.), представ-



ленных производителем (изготовителем) в испытательную медицинскую организацию;

- испытательная медицинская организация в обязательном порядке оценивает документацию на программный продукт, включая «Технические условия» (ТУ), и делает вывод как о том, что содержание документа доступно и понятно специалистам соответствующего профиля, так и о том, что его содержание обеспечивает надлежащее эффективное и безопасное медицинское применение.

Результаты клинических испытаний Программного обеспечения (как и любого другого медицинского изделия) должны быть оформлены в соответствии с Приложениями к приказу Министерства здравоохранения Российской Федерации от 09.01.2014 № 2н «Об утверждении Порядка проведения оценки соответствия медицинских изделий в форме технических испытаний, токсикологических исследований, клинических испытаний в целях государственной регистрации медицинских изделий».

При подготовке документации необходимо формировать подробные результаты оценки и анализа данных в соответствии с дополненным чек-листом STARD 2015 (см. Приложение).

ЗАКЛЮЧЕНИЕ

Отдельные алгоритмы и инструменты на основе интеллектуальных технологий уже успешно применяются в диагностике. Дальнейшая разработка «искусственного интеллекта» должна быть направлена на решение конкретных медицинских задач.

Внедрение данных рекомендаций позволит унифицировать проведение клинических испытаний программного обеспечения, разработанного на основе интеллектуальных технологий для сферы лучевой диагностики. Такие испытания – обязательный компонент для регистрации «искусственного интеллекта» в качестве медицинского изделия.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Власов В.В., Реброва О.Ю. Доказательная медицина: оценка эффективности диагностических вмешательств. Заместитель главного врача. 2010. № 4. С. 50.
2. Методические рекомендации по порядку проведения экспертизы качества, эффективности и безопасности медицинских изделий (в части программного обеспечения) для государственной регистрации в рамках национальной системы / М.: ФГ БУ «ВНИИИИМТ» Росздравнадзора, 2018. – 31 с.
3. Николенко С., Кадурич А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018.
4. О порядке проведения аудита качества исследований КТ, МРТ, ПЭТ-КТ, маммографии» (№38 от 29.12.2016). Методические рекомендации, утвержденные Департаментом здравоохранения г. Москвы / Гомболевский В.А., Ламзин М.С., Будакова Л.А., Буренчев Д.В., Гусева Е.Б., Ким С.Ю., Морозов С.П. – М., 2016. – 100 с.
5. Allen B, Dreyer K. The Role of the ACR Data Science Institute in Advancing Health Equity in Radiology. *J Am Coll Radiol*. 2019 Apr; 16(4 Pt B):644-648. DOI: 10.1016/j.jacr.2018.12.038.
6. Artificial Intelligence in Medical Imaging. Opportunities, Applications and Risks. Erik R. Ranschaert, Sergey Morozov, Paul R. Algra (Eds.); Springer, 2019.
7. Bossuyt PM, Reitsma JB, Bruns DE et al., For the STARD Group. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. *Radiology*. 2015;151:1516. PMID: 26509226.
8. Guidance for Industry and Food and Drug Administration Staff - Computer-Assisted Detection Devices Applied to Radiology Images and Radiology Device Data - Premarket Notification. <https://www.fda.gov/MedicalDevices/ucm187249.htm>.
9. RCR position statement on artificial intelligence. – 20.07.2018. <https://www.rcr.ac.uk/posts/rcr-position-statement-artificial-intelligence>.
10. Tang A, Tam R, Cadrin-Chênevert A et al. Canadian Association of Radiologists White Paper on Artificial Intelligence in Radiology. *Can Assoc Radiol J*. 2018 May;69(2):120-135. DOI: 10.1016/j.carj.2018.02.002.

Приложение 1

Критерии для допуска ИИ к предварительной опытной эксплуатации

1. Цели:

- ПО обеспечивает предварительный автоматический анализ диагностических изображений для повышения качества и скорости подготовки описания врачом;
- ПО обеспечивает приоритизацию исследований в рабочем списке («worklist») в соответствии с автоматически выявленной патологией;
- ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики);
- ПО обеспечивает поддержку принятия врачебных решений;
- ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ.

2. Сертификация:

- наличие у ПО на основе интеллектуальных технологий сертификации FDA и/или CE (класс II); наличие у компании-разработчика фактического внедрения ПО в медицинские организации: не менее 2-х независимых учреждений; более 6 месяцев эксплуатации; не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ПО решает несколько задач);
- наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых «Scopus» и/или «Web of Science» и входящих в первый-второй квартиль согласно «International Scientific Journal & Country Ranking»;
- должны быть доказаны диагностическая точность – $AUC > 0,8$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж);
- наличие или готовность пройти процедуру государственной регистрации в качестве медицинского изделия в РФ.

3. Безопасность:

- соблюдение требований законодательства Российской Федерации в сфере персональных данных, информационной безопасности, охраны здоровья;
- наличие или готовность разместить на территории Российской Федерации серверные мощности, необходимые для работы ПО.



4. Доказательность:

- после окончания разработки оценка точности алгоритмов проводилась на независимых данных ;
- диагностическая точность протестирована на популяции, характеристики которой аналогичны населению территории предполагаемого использования ИИ;
- ежегодное обновление информации о диагностической точности.

5. Стандартизация:

- автоматизированный анализ диагностических изображений в стандарте DICOM;
- поддержка стандарта HealthLevel7 (HL7) / FHIR (в частности должен обеспечиваться обмен сообщениями о завершении автоматического анализа изображения, обнаружении и классификации патологии);
- при формировании проектов описаний используются рекомендованные классификации (RADS, MAGNIMS и т.д.).

6. Интеграция:

- наличие или готовность разработать средства для «бесшовной» интеграции с информационными системами в сфере здравоохранения данного субъекта РФ, медицинскими информационными системами;
- наличие или готовность разработать средства для интеграции с радиологической информационной системой «AGFA Enterprise Imaging» или ее аналогами;
- наличие средств для интеграции с PACS и ПИС (DICOM «query and retrieve»);
- возможность «бесшовной» интеграции с PACS/ПИС, обеспечиваемой только программными средствами.

7. Функциональность:

- возможность потоковой обработки с последующей отправкой в PACS серий, дополненных результатами анализа ИИ;
- возможность совмещения серий нативных изображений и серий, дополненных результатами анализа ИИ;
- выявление находок (нозологий, патологических признаков, отклонений от нормы) для данной модальности или рабочего процесса в соответствии с техническим заданием;

²База медицинских данных для тестирования отличалась от базы, использованной для обучения, калибровки и валидации системы (то есть тестирование осуществлено на данных, с которыми ранее алгоритм не работал).

- классификация (определение типа, разновидности) находки, в том числе в соответствии с Международной классификацией болезней (версия 10) если предусмотрено техническим заданием;
- предоставление информации о вероятности наличия целевой патологии;
- автоматическая подготовка и выгрузка в PACS проекта описаний находок (с соблюдением структуры шаблона описаний, предустановленной в PACS);
- результат автоматического анализа отображается стандартными средствами PACS, в том числе графическими возможностями DICOM;
- наличие шаблонов протоколов описаний с их автоматической генерацией и отправкой в РИС/МИС посредством сообщений HL7/FHIR;
- наличие встроенного инструмента оценки точности;
- возможность дополнительной тренировки алгоритма на обрабатываемых («real world») данных;
- длительность обработки одного исследования не превышает 60 секунд, не учитывая время на пересылку данных. При оценке динамики (сравнении исследований) анализ может занимать более 60 секунд, но не более 60 секунд на одно исследование;
- функция контроля, обеспечивающая старт автоматического анализа только по факту получения из РИС сообщения HL7 «status complete by radiographer»;
- сообщения ПО в стандарте HL7 о завершении автоматического анализа и обнаружении патологии используются для приоритизации исследований в рабочем списке («worklist»);
- возможность автоматизированного поиска в базе PACS/РИС аналогичных исследований для сравнения;
- для каждого проанализированного исследования ПО формирует декларацию, содержащую название ПО (алгоритма), список выявленных находок (нозологий, патологических признаков, отклонений от нормы) с указанием чувствительности и специфичности для каждого типа находок; декларация формируется в DICOM SR, Overlay и т.д. и передается в PACS как отдельная серия;
- результатом работы ПО является сформированная серия изображений в формате DICOM, которая:
 - а) имеет количество срезов аналогичное исходной серии для возможности синхронного просмотра врачом;
 - б) обладает информацией на каждом срезе о названии ПО, версии, диагностической точности с датой ее проверки и точным временем завершеного анализа конкретного исследования;



с) допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок).

8. Контракт:

- наличие у компании-разработчика системы менеджмента качества;
- наличие у компании-разработчика политики подготовки и контроля версий;
- регулярные обновления системы, включая обновление информации о диагностической точности;
- обновления ПО включены в стоимость;
- все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика.

Анкета для допуска программного обеспечения (ПО) на основе технологий «искусственного интеллекта» / компьютерного зрения к предварительной опытной эксплуатации

Раздел	Метрики	Ответ (1-да/0-нет)	Комментарий
1. Цели	<p>1.1. ПО обеспечивает предварительный автоматический анализ медицинских изображений в формате DICOM для повышения качества и скорости подготовки описания врачом.</p> <p>1.2. ПО обеспечивает приоритизацию исследований в рабочем списке («worklist») в соответствии с автоматически выявленной патологией.</p> <p>1.3. ПО обеспечивает автоматическую подготовку проекта описания исследования по результатам работы алгоритма ИИ.</p> <p>1.4. ПО обеспечивает предварительный сравнительный анализ исследований одного пациента в формате DICOM, выполненных в разное время (оценка динамики).</p>		
2. Сертификация	<p>2.1. Наличие у ПО на основе интеллектуальных технологий сертификации FDA и/или CE (класс II).</p> <p>В случае ответа «нет» на пункт 2.1, должны быть ответы «да» на пункты 2.2 и 2.3:</p> <p>2.2. Наличие у компании-разработчика фактического внедрения ПО в медицинские организации:</p> <ul style="list-style-type: none"> • не менее 2-х независимых учреждений; • более 6 месяцев эксплуатации; • не менее 1000 успешно выполненных анализов (подтвержденных врачом-диагностом) для каждой задачи (если ПО решает несколько задач). 		



	<p>2.3. Наличие научных статей (оригинальных исследований) в рецензируемых журналах, индексируемых «Scopus» и/или «Web of Science» и входящих в первый-второй квартиль согласно «International Scientific Journal & Country Ranking»; должны быть доказаны диагностическая точность – $AUC \geq 0,8$ (классическая ROC-кривая) и повышение эффективности производственных процессов (сравнение работы специалистов с и без данного алгоритма ИИ, в том числе хронометраж).</p>		
3. Доказательность	<p>3.1. После окончания разработки оценка точности алгоритмов проводилась на независимых данных (база медицинских данных для тестирования отличалась от базы, использованной для обучения, калибровки и валидации системы; то есть тестирование осуществлено на данных, с которыми ранее алгоритм не сталкивался).</p> <p>3.2. Диагностическая точность протестирована на данных лиц европеоидной и монголоидной рас.</p> <p>3.3. Ежегодное обновление информации о диагностической точности.</p>		
4. Функциональность	<p>4.1. Наличие встроенного инструмента оценки точности.</p> <p>4.2. Длительность обработки одного исследования не превышает 60 секунд, не учитывая время на пересылку данных. Для выполнения цели 1.4 анализ может занимать более 60 секунд, но не более 60 сек на одно исследование.</p> <p>4.3. Результатом работы ПО является сформированная серия изображений в формате DICOM, которая:</p> <ul style="list-style-type: none">• имеет количество срезов аналогичное исходной серии для возможности синхронного просмотра врачом;		

	<ul style="list-style-type: none"> • обладает информацией на каждом срезе о названии ПО, версии, диагностической точности с датой ее проверки и точным временем завершеного анализа конкретного исследования. • допускается наличие дополнительных серий с информацией о результатах анализа (сводные таблицы анализа выявленных находок в динамике и/или отдельные изображения находок). 		
<p>5. Контракт</p>	<p>5.1. Регулярные обновления системы, включая обновление информации о диагностической точности.</p> <p>5.2. Обновления ПО включены в стоимость.</p> <p>5.3. Все медицинские данные, сопутствующие материалы и результаты работы ПО являются собственностью заказчика.</p>		



Приложение 2

Чек-лист для детального описания результатов клинических испытаний программного обеспечения, разработанного для лучевой диагностики на основе интеллектуальных технологий (на основе STARD 2015 [7] с дополнениями)

Раздел и тема	№	Пункт
НАЗВАНИЕ		
	1	Акцент на оценку диагностической точности программного обеспечения на основе интеллектуальных технологий, предназначенного для лучевой диагностики
РЕЗЮМЕ		
	2	Структурированное изложение дизайна исследования, материалов и методов, результатов и заключения
ВВЕДЕНИЕ		
	3	Клинические и научные сведения, включающие практическое применение результатов нового (исследуемого) метода – индекс-теста
	4	Гипотеза, цели и конечные точки исследования
МЕТОДЫ		
Дизайн исследования	5	Проспективный или ретроспективный характер исследования
Выборка	6	Критерии включения в исследование (включая целевую патологию)
	7	На основании каких критериев проведен отбор пациентов (симптомы, результаты обследований и т.д.)
	8	Где и когда сформирована база медицинских данных с указанием ключевых характеристик (государственная регистрация, характеристика популяции, сведения о депersonализации, целевая патология и диагностические группы с распределением, характеристика исследований, верификация, методология разметки)
	9	Запланированный размер выборки и способ его определения
	10a	Выборка сформирована последовательно, случайно или другими способами. Обоснование размера выборки.

Методы исследования	10b	Информация об индекс-тесте в достаточном для воспроизведения результатов объеме (включая сведения об использованных при разработке математических моделях, нейронных сетях, методах машинного обучения, выборках для обучения и калибровки)
	11	Информация о референс-тесте («золотом стандарте») в достаточном для воспроизведения результатов объеме (референс-тестом является разметка эталонных данных)
	12a	Обоснование выбора референс-теста (методологии разметки, всей базы эталонных данных)
	12b	Порог активации (точка отсечения, «cut-off») для индекс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации («cut-off»): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13a	Порог активации (порог решающего правила, «cut-off») для референс-теста, включая разделение исходов на «ожидаемые» и «непредвиденные». Метод определения порога активации: активации (cut-off): минимальное расстояние от верхнего левого угла до ROC-кривой, индекс Юдена, в зависимости от целей исследования
	13b	Была ли доступна участвующим в проведении/анализе индекс-теста критичная клинично-лабораторная информация, включая результаты применения референс-теста (было ли исследование слепыми или нет, во втором случае – обоснование причины)
	14	Была ли доступна участвующим в разметке и анализе эталонных данных критичная клинично-лабораторная информация, включая результаты применения индекс-тест (было ли исследование слепыми или нет, во втором случае – обоснование причины)
Анализ	15	Методы оценки и сравнения диагностической точности (метрики)
	16	Тактика при неоднозначных результатах индекс-теста или референс-теста
	17	Тактика при потере данных индекс-теста или референс-теста
	18	Анализ причин вариабельности диагностической точности с разделением на «ожидаемые» и «непредвиденные»



РЕЗУЛЬТАТЫ		
Пациенты	19	Характеристика пациентов: диспозиция пациентов, определение популяции для анализа, число пациентов, включенных в исследование, число досрочно выбывших, число пропущенных или негодных данных. Рекомендуется в дополнение к текстовому описанию представить диаграмму CONSORT
	20	Демографические и клинические характеристики выборки на момент первичного обследования (всех субъектов, а также субъектов, у которых есть результаты обоих тестов)
	21a	Распределение по степени тяжести целевой патологии
	21b	Распределение прочих диагнозов (по Международной классификации болезней) у пациентов без целевой патологии
	22	Любые значимые различия в методиках проведения индекс-теста и референс-теста
Результаты	23	Комбинированная таблица результатов индекс-теста и референс-теста
	24	Подсчет метрик диагностической точности с определенным доверительным интервалом (например, 95%)
	25	Любые побочные эффекты вследствие индекс-теста или референс-теста
ОБСУЖДЕНИЕ		
	26	Ограничения исследования, включая источники систематических ошибок, статистических неточностей и обобщений
	27	Клиническая ценность, включая планируемое применение индекс-теста
ПРОЧЕЕ		
	28	Сведения о регистрации исследования (любого этапа)
	29	Где можно получить полные сведения о разработке (включая информацию о базе медицинских данных для обучения и калибровки)
	30	Источники финансирования и прочее