

Оценка надежности измерений легочных узлов при компьютерной томографии: ретроспективное исследование

Ю.А. Васильев¹, И.А. Блохин^{1✉}, BlokhinIA@zdrav.mos.ru, М.Р. Коденко^{1,2}, М.М. Сучилова¹, А.В. Владзимирский^{1,3}, О.В. Омелянская¹, Р.В. Решетников¹

¹ Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1

² Московский государственный технический университет имени Н.Э. Баумана; 105005, Россия, Москва, ул. 2-я Бауманская, д. 5

³ Первый Московский государственный медицинский университет имени И.М. Сеченова (Сеченовский Университет); 119048, Россия, Москва, ул. Трубецкая, д. 8, стр. 2

Резюме

Введение. В XIX веке медицинская визуализация практически повсеместно перешла от простой оценки наличия патологического признака к количественной оценке выявленных изменений. Несмотря на явную зависимость диагностической точности от степени погрешности при выполнении измерений, современные классификации рентгенологических ошибок не выделяют погрешность измерения как отдельную причину расхождений между специалистами. Мы считаем необходимым подробно изучить размер погрешности измерений врачей-рентгенологов в условиях, приближенных к рутинной практике, и определить возможность влияния такой ошибки на формирование правильной тактики лечения пациента.

Цель. Определить погрешность врачей-рентгенологов при измерении размеров легочного узла в ходе оппортунистического скрининга при компьютерной томографии органов грудной клетки (КТ ОГК).

Материалы и методы. Ретроспективное исследование проведено на материалах базы данных, содержащей разметку 3 897 нативных исследований КТ ОГК по стандартному протоколу. Разметка включала определение типа и среднего размера каждого легочного узла. Использовались следующие статистические методы обработки данных: оценка согласия разметчиков в оценке размеров очага – внутриклассовый коэффициент корреляции, погрешность измерений – пределы согласия (метод Бланда – Альтмана). Уровень значимости для всех статистических гипотез был принят равным 0,05.

Результаты. В итоговый анализ после исследования на наличие выбросов включено 87 КТ-исследований, каждое из которых содержало единственный солидный легочный узел диаметром от 6 до 30 мм, независимо измеренный тремя врачами-рентгенологами. Медианный размер узла составил 10 мм. Внутриклассовый коэффициент корреляции составил 0,95 при 95% ДИ (0,93; 0,97). 95%-ный ДИ для максимальной абсолютной разности между разметчиками находится в диапазоне (1,5; 1,9) мм.

Выводы. Погрешность измерения размеров солидного легочного узла при КТ ОГК тремя врачами-рентгенологами не превышает методологически установленного порога роста легочного узла.

Ключевые слова: компьютерная томография, легочный очаг, надежность измерений

Благодарности. Данная статья подготовлена авторским коллективом в рамках научно-практического проекта в сфере медицины (№ ЕГИСУ: 123031500005-2) «Научное обоснование методов лучевой диагностики опухолевых заболеваний с использованием радиомического анализа».

Для цитирования: Васильев ЮА, Блохин ИА, Коденко МР, Сучилова ММ, Владзимирский АВ, Омелянская ОВ, Решетников РВ. Оценка надежности измерений легочных узлов при компьютерной томографии: ретроспективное исследование.

Медицинский совет. 2025;19(21):123–129. <https://doi.org/10.21518/ms2025-522>.

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Lung nodule measurement reliability in computed tomography: A retrospective study

Yuriy A. Vasilev¹, Ivan A. Blokhin^{1✉}, BlokhinIA@zdrav.mos.ru, Maria R. Kodenko^{1,2}, Maria M. Suchilova¹, Anton V. Vladzimyrskyy^{1,3}, Olga V. Omelyanskaya¹, Roman V. Reshetnikov¹

¹ Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia

² Bauman Moscow State Technical University; 5, 2nd Baumanskaya St., Moscow, 105005, Russia

³ Sechenov First Moscow State Medical University (Sechenov University); 8, Bldg. 2, Trubetskaya St., Moscow, 119991, Russia

Abstract

Introduction. In the 21st century, medical imaging evolved from simple assessment of the presence of a pathological sign to quantitative assessment of detected changes. Despite the obvious dependence of diagnostic accuracy on the degree of measurement error, modern classifications of radiologic errors do not single out measurement error as a separate cause of discrepan-

cies between specialists. We consider it necessary to study in detail the size of measurement error of radiologists in conditions close to routine practice and to determine the possibility of the influence of such an error on the formation of the correct patient management tactics.

Aim. To determine radiologist discrepancy in measuring the pulmonary nodule size during opportunistic screening in chest computed tomography (CT).

Materials and methods. A retrospective study was performed on a database containing the 3897 labeled native chest CT studies using a routine protocol. Experts determined the type and mean size of each pulmonary nodule. We used the following statistical methods: intraclass correlation coefficient to evaluate measurement reliability; limits of agreement (Bland-Altman method) to quantify measurement variability. The significance level for all statistical hypotheses was set to 0.05.

Results. The final analysis included 87 CT studies, each containing a single solid pulmonary nodule with a diameter ranging from 6 to 30 mm, independently measured by three radiologists. The median nodule size was 10 mm. The intraclass correlation coefficient was 0.95 with 95% CI (0.93; 0.97). The 95% CI for the maximum absolute difference between observers ranged between (1.5; 1.9) mm.

Conclusion. The measurement discrepancy between three radiologists for solid pulmonary nodule in chest CT did not exceed the methodologically established threshold for pulmonary nodule growth.

Keywords: X-Ray Computed Tomography, Solitary Lung Nodule, Finding Reproducibility

Acknowledgments. This article was prepared by a group of authors as a part of the medical research project (No. USIS [in the Unified State Information System for Accounting of Research, Development, and Technological Works]: 123031500005-2) "Scientific rationale for diagnostic radiology in oncology using radiomics".

For citation: Vasilev YuA, Blokhin IA, Kodenko MR, Suchilova MM, Vladzemyrskyy AV, Omelyanskaya OV, Reshetnikov RV. Lung nodule measurement reliability in computed tomography: A retrospective study. *Meditinskiy Sovet*. 2025;19(21):123–129. (In Russ.) <https://doi.org/10.21518/ms2025-522>.

Conflict of interest: the authors declare no conflict of interest.

ВВЕДЕНИЕ

В XIX в. медицинская визуализация практически по-всеместно перешла от простой оценки наличия патологического признака к количественной оценке выявленных изменений [1]. Такой переход позволил стандартизировать подходы к оценке патологии и широко внедрить системы классификации заболеваний для оценки риска и выбора дальнейшей тактики лечения.

Наличие количественных характеристик патологий в современных системах описания результатов лучевой диагностики (reporting system) приводит к тому, что метрики их диагностической точности становятся зависимыми от степени погрешности при выполнении измерений. Например, при диагностике аневризмы брюшной аорты методом компьютерно-томографической ангиографии погрешность измерений диаметра между врачами-рентгенологами может составлять более пяти миллиметров, что необходимо учитывать при принятии врачебных решений [2]. Несмотря на подобное влияние, современные классификации рентгенологических ошибок не выделяют погрешность измерения как отдельную причину расхождений между специалистами в повседневной практике и при аудите.

Измерение среднего диаметра, наряду с определением типа легочного узла, лежит в основе их оценки при компьютерной томографии органов грудной клетки (КТ ОГК) [3] для выявления признаков, коррелирующих со злокачественным новообразованием легких. По терминологии общества Fleishner (Fleishner Society), адаптированной на территории Российской Федерации, при КТ ОГК узел определяется как уплотнение сферической или

неправильной формы, диаметром до 3 см [4]. Тактика ведения выявленного узла определяется не только его типом и размером, но и скоростью роста. Детекция этой динамики зависит от погрешности при проведении измерений диаметра очага, которая позволяет достоверно зафиксировать только рост, превышающий некоторое пороговое значение. При интерпретации НДКТ с использованием Lung-RADS принято указывать рост узла, если наблюдается увеличение его диаметра на 1,5 мм за 12 мес.¹, а при оценке КТ по критериям общества Fleishner – на 2 мм [5]. Отметим, что первая цифра метрологически не обоснована [6], а вторая базируется на исследовании 2004 г., выполненном в контролируемых лабораторных условиях, которые сложно воспроизвести в клинической практике [7]. В связи с этим мы считаем необходимой актуализацию погрешности измерений врачей-рентгенологов в условиях, приближенных к рутинной практике.

Цель исследования – определить ошибку врачей-рентгенологов при измерении размеров легочного узла на КТ ОГК. Гипотеза исследования: погрешность измерения среднего диаметра солидного легочного узла, рассчитанная для трех независимых экспертов, не превышает 2 мм.

МАТЕРИАЛЫ И МЕТОДЫ

Настоящее ретроспективное исследование проведено на материалах базы данных (БД) [8], полученной в процессе работы проекта «Московский скрининг рака легкого» [9, 10]. БД содержит разметку 3 897 уникальных

¹ Lung Rads. American College of Radiology n.d. Available at: <https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Lung-Rads>.

нативных исследований КТ ОГК. В анализ были включены исследования, содержащие легочные очаги, независимо идентифицированные и размеченные тремя врачами-рентгенологами с опытом работы от 2 до 10 лет. Параметры сканирования: зона сканирования от верхней границы ключицы через оба реберно-диафрагмальных угла, 120 kV, 110 mAs, питч 1,3, толщина среза 2 мм, реконструкции с мягкотканным (B30f) и легочным (B80f) фильтрами реконструкции (convolution kernel). Разметка исследований включала определение типа и среднего линейного размера каждого легочного узла в аксиальной плоскости с округлением до ближайшего целого числа. Критерии включения и их обоснование приведены в таблице. Разметка исследований осуществлялась при помощи инструмента FanToM (Find Anomalies in Tomography) [11] с возможностью удаленного доступа, что выгодно отличает его от других современных средств разметки медицинских изображений [12]. Вследствие этого эксперты пользовались широким набором конфигураций рабочих компьютеров с индивидуальными настройками яркости, контраста и гамма-коррекции, что позволяет моделировать разнообразие условий рутинной клинической практики. Настоящее исследование проведено в соответствии с критериями Guidelines for Reporting Reliability and Agreement Studies (GRRAS) [13] и методологией Abu-Arafeh et al. для анализа Бланда – Альтмана [14].

Расчет размера выборки

Расчет размера выборки основан на методе, описанном в книге Shein-Chung Chow et al. [15] для выполнения одновыборочного теста на эквивалентность (one-sample equivalence test). В контексте работы предполагается повторяемая независимая оценка размера каждого узла различными разметчиками. Нулевая гипотеза теста постулирует, что истинное отличие средних значений со-поставляемых выборок превышает допустимый порог.

Таблица. Критерии включения в исследование и их обоснование

Table. Criteria for inclusion in the study and their justification

Критерий включения	Обоснование
Наличие исследования в ЕРИС	Доступ к протоколу описания (при необходимости)
Не более одного очага в исследовании	При любой методологии скрининга финальная оценка выставляется по наиболее крупному очагу
Средний размер очага от 6 до 30 мм	Нижний порог очага, требующего активного наблюдения; верхний порог закреплен терминологически, находке более 30 мм присваивается наименование «образование» [4]
Солидный тип очага	Ограничение базы данных; не было предусмотрено отдельного измерения солидного компонента
Отсутствие дополнительных комментариев	Исключение исследований с тяжелой сопутствующей патологией, способной искажить результаты измерений

Для оценки минимально необходимого размера выборки были заданы уровень значимости ($\alpha = 0,05$), мощность теста ($1-\beta = 90\%$), граничное значение расхождения измерений (2 мм), ожидаемое расхождение (не более 1 мм) и оценка среднеквадратического отклонения расхождения (~ 2 мм). Также для учета возможного наличия исключений выборки (до 20%) размер выборки был масштабирован в 1,2 раза. Соответствующий размер выборки, согласно расчетам по предложенному методу, составляет 44 экспериментальные точки.

Статистический анализ

Общая выборка измеренных диаметров была проанализирована на предмет соответствия распределения нормальному с помощью теста Шапиро – Уилка. Далее был проведен анализ выборки для определения основных показателей описательной статистики: была построена гистограмма частот размеров в выборке, вычислены среднее значение диаметра очага и среднеквадратичное отклонение (в случае нормально распределенных данных) либо медиана и интерквартильный размах.

Для оценки погрешности измерений между экспертами был проведен расчет 95%-ного доверительного интервала (95% ДИ) для средней разности оценок размера узла. Визуализация пределов согласия была получена с помощью анализа Бланда – Альтмана (Bland-Altman). В силу того что дизайн исследования предполагает оценку одного узла более чем двумя экспертами, был использован подход расширенного анализа, предложенный M. Jones et al. [16]. Выбранная методика имеет ряд отличий от стандартного подхода. Вместо оценки средней разницы между двумя разметчиками оценивают их отклонение от усредненного значения измеряемой величины. Рассчитанные пределы согласия показывают, насколько измерение отдельного наблюдателя может отличаться от среднего значения всех наблюдателей. Вычисления были проведены с помощью функции BApplotMultipleR пакета AGREL для R. Результат применения функции объединяет следующие показатели: значение внутриклассового коэффициента корреляции, абсолютные значения пределов согласия и результаты дисперсионного анализа. Внутриклассовый коэффициент корреляции рассчитан на основе разброса размеров оцениваемого очага, его значение позволяет сделать вывод о согласованности разметчиков при оценке каждого исследуемого узла.

Уровень значимости для всех статистических гипотез был принят равным 0,05. Обработка данных выполнена с использованием языка R в программе RStudio Desktop версии 2023.03.1+446 [17].

РЕЗУЛЬТАТЫ

В итоговый анализ было включено 103 КТ-исследования, каждое из которых содержало единственный легочный узел диаметром от 6 до 30 мм, измеренный тремя врачами-рентгенологами, не отметившими в исследовании иной патологии. Данная выборка была дополнительно исследована на наличие выбросов по методу отсечения

значений, находящихся вне диапазона $[Q1-1,7*IQR; Q3+1,7*IQR]$, где $Q1$ и $Q3$ – первый и третий квартилы выборки, IQR – интерквартильный размах. Данный метод позволяет ограничить размер выборки областью, соответствующей трем СКО². Значения вне данного диапазона, т. н. «выбросы», были дополнительно просмотрены врачом-рентгенологом с опытом более 5 лет для проверки наличия методологических ошибок при измерении очага исходным разметчиком. Результат данного анализа позволил исключить 16 некорректно размеченных исследований. Таким образом, итоговая выборка включала 87 исследований, что двукратно превышает рассчитанный порог в 44 минимально необходимых случая. Блок-схема проведения исследования представлена на *рис. 1*.

Выборка размеров очагов имеет отличное от нормального распределение ($p\text{-value} < 0,001$). Медиана выборки составила 10 мм при интерквартильном размахе [7,6; 13] (*рис. 2A*). Максимальное значение абсолютной разности оценок разметчиков, иллюстрирующее «наихудший» сценарий расхождения, демонстрирует разброс от 0 до 4 мм при среднем значении 1,8 мм, причем большая часть значений сгруппирована между 1 и 2 мм (*рис. 2C*).

Результаты корреляционного анализа для усредненных значений размера очага по оценкам разметчиков и отклонения от данного размера демонстрируют отсутствие статистически значимой взаимосвязи (ρ Спирмена – 0,1, $p\text{-value} = 0,3$). Следовательно, рассчитанные пределы согласия для исследуемой выборки не зависят от размера обнаруженного очага, и анализ подгрупп нецелесообразен. Надежность результатов разметки подтверждается высоким значением коэффициента внутрикласовой корреляции (0,95 95% ДИ (0,93; 0,97)), соответствующим «отличному» согласию.

Расчет 95% ДИ для максимальной абсолютной разности между разметчиками демонстрирует диапазон (1,5; 1,9) мм, т. е. при худшем сценарии разброс оценок будет находится в диапазоне $\pm 1,9$ мм. Полученные результаты хорошо согласуются с результатами анализа Бланда – Альтмана, построенного с учетом всех значений, входящих в исходную выборку.

ОБСУЖДЕНИЕ

Настоящее ретроспективное исследование погрешности измерения легочных узлов при КТ ОГК с достаточным объемом выборки продемонстрировало 95% ДИ максимальной абсолютной разности между разметчиками, равной (1,5; 1,9) мм, и внутрикласовый коэффициент корреляции, соответствующий «отличному» согласию [18]. Гипотеза исследования была принята: на выборке из

Рисунок 1. Блок-схема формирования выборки легочных очагов в исследовании

Figure 1. Lung nodule sample flow chart

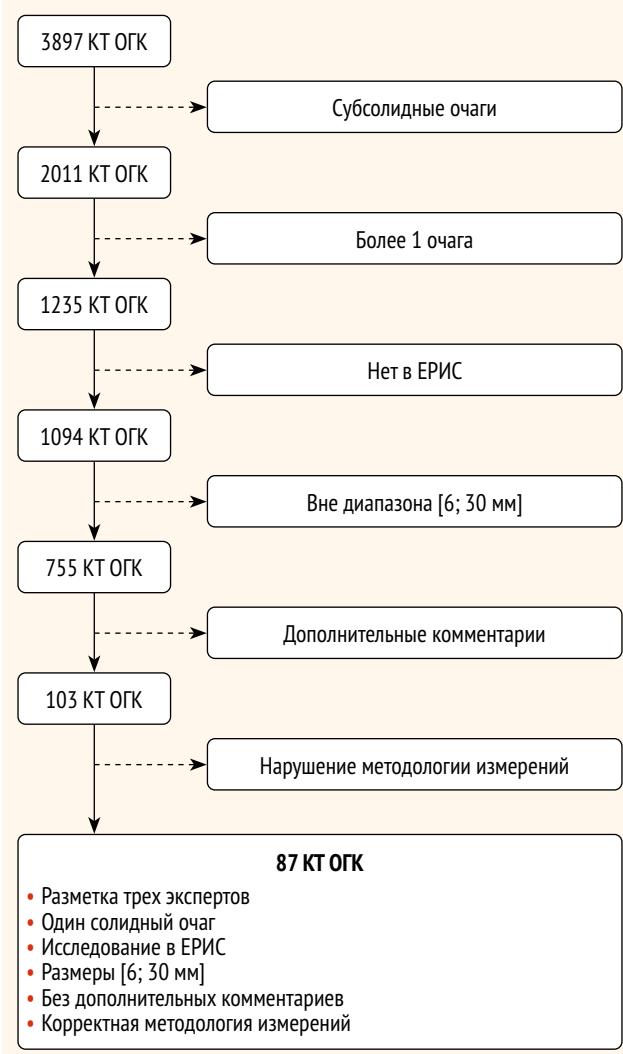
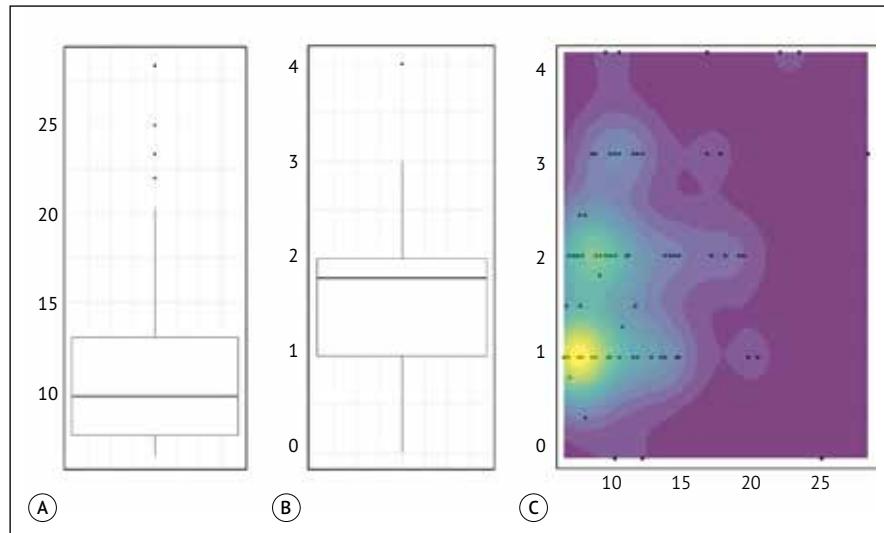


Рисунок 2. Описательная статистика выборки

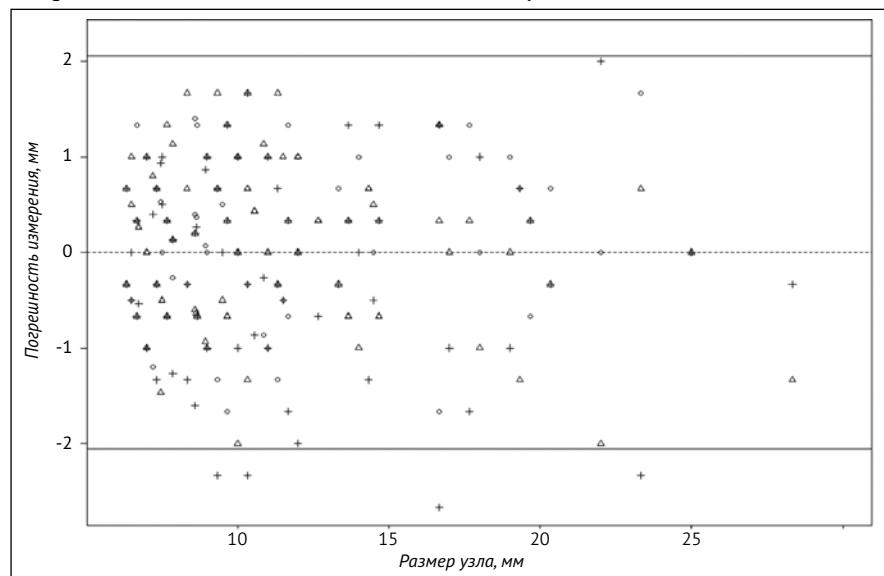
Figure 2. Descriptive statistics of the sample



Примечание: А – диаграмма распределения размера очага (усредненное значение по трем оценкам), мм; В – диаграмма максимальных значений разности оценок размера узла, мм; С – диаграмма зависимости погрешности измерения от размера узла

² Unmasking the Outliers: Exploring the Interquartile Range Method for Reliable Data Analysis ProCogia n.d. Available at: <https://procogia.com/interquartile-range-method-for-reliable-data-analysis>.

Рисунок 3. Результаты расширенного анализа Бланда – Альтмана
Figure 3. Results of extended Bland-Altman analysis



Примечание: треугольниками и крестами отмечены 1, 2 и 3-я оценка размера очага, горизонтальными прямыми отмечены границы согласия (мм)

87 легочных узлов, независимо размеченных тремя врачами-рентгенологами, полученные значения погрешности измерения не превышают 2 мм, т. е. границу роста легочного узла, заложенную в рекомендации общества Fleishner для интерпретации КТ ОГК [19].

Наши результаты согласуются с результатами других исследований. Например, в исследовании M.-P. Revel et al. 2004 г. при измерении легочных узлов диаметром до 2 см пределы согласия между тремя экспертами составили 1,73 мм [7]. K. Marten et al. продемонстрировали пределы согласия от -1,46 до 1,92 мм для легочных узлов [20]. Другое исследование, выполненное B. Zhao et al. в 2009 г., было посвящено определению погрешности измерений одного, двух размеров и объема узлов при двух последовательных КТ ОГК [21]. Пределы согласия составили от -7,3% до 6,2%, от -17,6% до 19,8% и от -12,1% до 13,4% соответственно. К сожалению, авторы приводят только относительную погрешность измерения в процентах.

Научная новизна настоящей работы обусловлена тем, что в исследовании M.-P. Revel et al. измерения проводились на экранах с разрешением 1280×1024 пикселей [7], а B. Zhao et al. не указывают, на каких мониторах производились измерения [21]. Настоящая работа проведена с использованием широкого спектра конфигураций современного оборудования. Представляется возможным, что контрастность и разрешение современных мониторов способны повлиять на точность измерений при интерпретации исследований [22, 23].

Причины ошибок в рентгенологии и их классификации одним из первых изучил L.H. Garland в 1949 г. на примере пленочной рентгенографии [24]. За последние десять лет предложено несколько классификаций врачебных ошибок в лучевой диагностике. S. Waite et al. [25] выделили внутренние и внешние источники ошибок интерпретации исследований. К первым относятся процессы выявления, распознавания, трактовки изменений, а также

когнитивные искажения и феномен удовлетворенного поиска. Ко вторым относятся отсутствие анамнеза пациента при описании, усталость эксперта, значительное количество исследований и отвлекающие факторы. A.P. Brady et al. приводят классификацию по категории пропуска патологии, удовлетворенного поиска, неправильного обоснования, находок вне области интереса, избыточной опоры на предыдущий протокол, трактовки в отрыве от предыдущего исследования, трактовки в отрыве от анамнеза и плохой коммуникации [26]. А.А. Учеваткин и соавт. проанализировали эпидемиологию и классификацию ошибок в лучевой диагностике, составив список вопросов для самоанализа, позволяющий минимизировать диагностические ошибки [27].

Все вышеперечисленные варианты классификации напрямую не выделяют ошибку, связанную с погрешностью измерения. Можно предположить, что это связано с недостаточным использованием количественных признаков для трактовки значительного числа патологических состояний. Например, из предложенных Американским колледжем радиологов систем классификации RADS (Reporting and Data System) в BI-RADS, NI-RADS и O-RADS, используются только качественные и полукачественные признаки (контуры, форма, сигнальные характеристики, кривая накопления контраста, относительный уровень метаболической активности и т. д.) [28]. С другой стороны, в системе оценки ответа солидных опухолей на проводимое лечение RECIST применяется сумма диаметров таргетных очагов, поэтому на итоговую оценку может повлиять именно погрешность измерения [29, 30]. Мы считаем необходимым выделить ошибку измерения в отдельный тип возможной ошибки врача-рентгенолога, как это было предложено для ультразвукового исследования плода [31].

К ограничениям настоящего исследования относятся: ретроспективный характер анализа, включение только одного типа очага, использование среднего диаметра с округлением до целого числа и невозможность сбора информации о конкретных конфигурациях и настройках рабочих машин разметчиков на момент выполнения измерений.

ВЫВОДЫ

Погрешность измерения размеров солидного легочного узла при компьютерной томографии органов грудной клетки тремя врачами-рентгенологами в ходе оппортунистического скрининга не превышает методологически установленного порога роста легочного узла.

Поступила / Received 29.01.2025
Поступила после рецензирования / Revised 01.07.2025
Принята в печать / Accepted 23.10.2025

Список литературы / References

1. Sullivan DC, Bresolin L, Seto B, Obuchowski NA, Raunig DL, Kessler LG. Introduction to metrology series. *Stat Methods Med Res.* 2015;24(1):3–8. <https://doi.org/10.1177/0962280214537332>.
2. Lederle FA, Wilson SE, Johnson GR, Reinke DB, Littooy FN, Acher CW et al. Variability in measurement of abdominal aortic aneurysms. *J Vasc Surg.* 1995;21(6):945–952. [https://doi.org/10.1016/S0741-5214\(95\)70222-9](https://doi.org/10.1016/S0741-5214(95)70222-9).
3. Larici AR, Farchione A, Franchi P, Ciliberto M, Cicchetti G, Calandriello L et al. Lung nodules: size still matters. *Eur Respir Rev.* 2017;26:170025. <https://doi.org/10.1183/16000617.0025-2017>.
4. Николаев АЕ, Сучилова ММ, Коркунова ОА, Блохин ИА, Гончар АП, Решетников РВ, Морозов СП. Терминология описания органов грудной клетки – рентгенография и компьютерная томография. М.: ГБУЗ «НПКЦ ДИТ ДЗМ»; 2022. 84 с.
5. Bankier AA, MacMahon H, Goo JM, Rubin GD, Schaefer-Prokop CM, Naidich DP. Recommendations for Measuring Pulmonary Nodules at CT: A Statement from the Fleischner Society. *Radiology.* 2017;285(2):584–600. <https://doi.org/10.1148/radiol.2017162894>.
6. Chelala L, Hossain R, Kazerooni EA, Christensen JD, Dyer DS, White CS. Lung-RADS Version 1.1: Challenges and a Look Ahead, From the AJR Special Series on Radiology Reporting and Data Systems. *Am J Roentgenol.* 2021;216(6):1411–1422. <https://doi.org/10.2214/AJR.20.24807>.
7. Revel M-P, Bissery A, Bienvenu M, Aycard L, Lefort C, Frija G. Are Two-dimensional CT Measurements of Small Noncalcified Pulmonary Nodules Reliable? *Radiology.* 2004;231(2):453–458. <https://doi.org/10.1148/radiol.2312030167>.
8. Морозов СП, Кульберг НС, Гомболовский ВА, Ледихова НВ, Сокolina ИА, Владимировский АВ, Бардин АС. Свидетельство о государственной регистрации базы данных RU 2018620500. Тегированные результаты компьютерных томографий легких. 2018. Режим доступа: <https://www.elibrary.ru/item.asp?id=39294439>.
9. Morozov SP, Kuzmina ES, Vetsheva NN, Gombolevskiy VA, Lantukh ZA, Polishuk NS et al. Moscow Screening: Lung Cancer Screening With Low-Dose Computed Tomography. *Probl Sotsialnoi Gig Zdravookhranenniia Istor Med.* 2019;27:630–636. <https://doi.org/10.32687/0869-866X-2019-27-si1-630-636>.
10. Владимировский АВ, Новик ВП, Павлов НА, Андрейченко АЕ, Блохин ИА, Гомболовский ВА и др. Свидетельство о государственной регистрации базы данных RU 2023621103. MosMedData: КТ с признаками рака легкого. 2023. Режим доступа: <https://elibrary.ru/item.asp?id=52121950>.
11. Morozov SP, Gombolevskiy VA, Elizarov AB, Gusev MA, Novik VP, Prokudaylo SB et al. A simplified cluster model and a tool adapted for collaborative labeling of lung cancer CT scans. *Comput Methods Programs Biomed.* 2021;206:106111. <https://doi.org/10.1016/j.cmpb.2021.106111>.
12. Васильев ЮА, Савкина ЕФ, Владимировский АВ, Омелянская ОВ, Арзамасов КМ. Обзор современных средств разметки цифровых диагностических изображений. Казанский медицинский журнал. 2023;104(5):750–760. <https://doi.org/10.17816/KMJ349060>.
Vasilev YA, Savkina EF, Vladzimyrskyy AV, Omelyanskaya OV, Arzamasov KM. Overview of modern digital diagnostic image markup tools. *Kazan Medical Journal.* 2023;104(5):750–760. (In Russ.) <https://doi.org/10.17816/KMJ349060>.
13. Kottner J, Audigé L, Brorson S, Donner A, Gajewski BJ, Hróbjartsson A et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96–106. <https://doi.org/10.1016/j.jclinepi.2010.03.002>.
14. Abu-Arafeh A, Jordan H, Drummond G. Reporting of method comparison studies: a review of advice, an assessment of current practice, and specific suggestions for future reports. *Br J Anaesth.* 2016;117(5):569–575. <https://doi.org/10.1093/bja/aew320>.
15. Chow S-C, Shao J, Wang H, Lokhnygina Y (eds.). *Sample size calculations in clinical research.* Boca Raton: Taylor & Francis; 2017.
16. Jones M, Dobson A, O'Brian S. A graphical method for assessing agreement with the mean between multiple observers using continuous measures. *Int J Epidemiol.* 2011;40(5):1308–1313. <https://doi.org/10.1093/ije/dyr109>.
17. Блохин ИА, Коденко МР, Шумская ЮФ, Гончар АП, Решетников РВ. Проверка гипотез исследования с использованием языка R. *Digital Diagnostics.* 2023;4(2):238–247. <https://doi.org/10.17816/DD121368>.
Blokhin IA, Kodenko MR, Shumskaya YF, Gonchar AP, Reshetnikov RV. Hypothesis testing using R. *Digital Diagnostics.* 2023;4(2):238–247. (In Russ.) <https://doi.org/10.17816/DD121368>.
18. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med.* 2016;15(2):155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>.
19. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology.* 2017;284(1):228–243. <https://doi.org/10.1148/radiol.2017161659>.
20. Marten K, Auer F, Schmidt S, Kohl G, Rummey EJ, Engelke C. Inadequacy of manual measurements compared to automated CT volumetry in assessment of treatment response of pulmonary metastases using RECIST criteria. *Eur Radiol.* 2006;16(4):781–790. <https://doi.org/10.1007/s00330-005-0036-x>.
21. Zhao B, James LP, Moskowitz CS, Guo P, Ginsberg MS, Lefkowitz RA et al. Evaluating Variability in Tumor Measurements from Same-day Repeat CT Scans of Patients with Non-Small Cell Lung Cancer. *Radiology.* 2009;252(1):263–272. <https://doi.org/10.1148/radiol.2522081593>.
22. Ziefle M. Effects of Display Resolution on Visual Performance. *Hum Factors.* 1998;40(4):554–568. <https://doi.org/10.1518/001872098779649355>.
23. Al-Ekrish AA, Ekram MIH, Al-Faleh W, AlKhader M, Al-Sadhan R. The validity of different display monitors in the assessment of dental implant site dimensions in cone beam computed tomography images. *Acta Odontol Scand.* 2013;71(5):1085–1091. <https://doi.org/10.3109/00016357.2012.741709>.
24. Garland LH. On the Scientific Evaluation of Diagnostic Procedures: Presidential Address Thirty-fourth Annual Meeting of the Radiological Society of North America. *Radiology.* 1949;52:309–328. <https://doi.org/10.1148/52.3.309>.
25. Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive Error in Radiology. *AJR Am J Roentgenol.* 2017;208(4):739–749. <https://doi.org/10.2214/AJR.16.16963>.
26. Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights Imaging.* 2017;8(1):171–182. <https://doi.org/10.1007/s13244-016-0534-1>.
27. Учеваткин АА, Юдин АЛ, Афанасьевна НИ, Юматова ЕА. Оттенки серого: как и почему мы ошибаемся. *Медицинская визуализация.* 2020;24(3):123–145. <https://doi.org/10.24835/1607-0763-2020-3-123-145>.
Uchevatin AA, Yudin AL, Afanas'yeva NI, Yumatova EA. Shades of grey: how and why we make mistakes. *Medical Visualization.* 2020;24(3):123–145. (In Russ.) <https://doi.org/10.24835/1607-0763-2020-3-123-145>.
28. Aras M, Erdil TY, Dane F, Gungor S, Ones T, Dede F et al. Comparison of WHO, RECIST 1.1, EORTC, and PERCIST criteria in the evaluation of treatment response in malignant solid tumors. *Nucl Med Commun.* 2016;37(1):9–15. <https://doi.org/10.1097/MNM.0000000000000401>.
29. Schwartz LH, Litière S, De Vries E, Ford R, Gwyther S, Mandrekar S et al. RECIST 1.1 –Update and clarification: From the RECIST committee. *Eur J Cancer.* 2016;62:132–137. <https://doi.org/10.1016/j.ejca.2016.03.081>.
30. Yoon SH, Kim KW, Goo JM, Kim D-W, Hahn S. Observer variability in RECIST-based tumor burden measurements: a meta-analysis. *Eur J Cancer.* 2016;53:5–15. <https://doi.org/10.1016/j.ejca.2015.10.014>.
31. Pinto A. Spectrum of diagnostic errors in radiology. *WJR.* 2010;2(10):377. <https://doi.org/10.4329/wjr.v2.i10.377>.

Вклад авторов:

Концепция статьи – Ю.А. Васильев, А.В. Владзимирский, О.В. Омелянская

Написание текста – И.А. Блохин, М.Р. Коденко, М.М. Сучилова

Сбор и обработка материала – И.А. Блохин, Р.В. Решетников

Обзор литературы – М.М. Сучилова

Статистическая обработка – М.Р. Коденко, Р.В. Решетников

Редактирование – М.М. Сучилова

Утверждение окончательного варианта статьи – Ю.А. Васильев, А.В. Владзимирский, О.В. Омелянская

Contribution of the authors:

Concept of the article – Yuriy A. Vasilev, Anton V. Vladzimyrskyy, Olga V. Omelyanskaya

Text development – Ivan A. Blokhin, Maria R. Kodenko, Maria M. Suchilova

Collection and processing of material – Ivan A. Blokhin, Roman V. Reshetnikov

Literature review – Maria R. Kodenko, Maria M. Suchilova

Statistical processing – Maria R. Kodenko, Roman V. Reshetnikov

Editing – Maria M. Suchilova

Approval of the final version of the article – Yuriy A. Vasilev, Anton V. Vladzimyrskyy, Olga V. Omelyanskaya

Информация об авторах:

Васильев Юрий Александрович, д.м.н., директор, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; <https://orcid.org/0000-0002-5283-5961>; npcmr@zdrav.mos.ru

Блохин Иван Андреевич, к.м.н., начальник сектора исследований в лучевой диагностике отдела научных медицинских исследований, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; <https://orcid.org/0000-0002-2681-9378>; BlokhinIA@zdrav.mos.ru

Коденко Мария Романовна, к.т.н., младший научный сотрудник отдела научных медицинских исследований, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; ассистент кафедры «Биомедицинские технические системы», Московский государственный технический университет имени Н.Э. Баумана; 105005, Россия, Москва, ул. 2-я Бауманская, д. 5; <https://orcid.org/0000-0002-0166-3768>; KodenkoMR@zdrav.mos.ru

Суцилова Мария Максимовна, младший научный сотрудник отдела научных медицинских исследований, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; <https://orcid.org/0000-0003-1117-0294>; SuchilovaMM@zdrav.mos.ru

Владзимирский Антон Вячеславович, д.м.н., заместитель директора по научной работе, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; профессор кафедры информационных технологий и обработки медицинских данных, Первый Московский государственный медицинский университет имени И.М. Сеченова (Сеченовский Университет); 119048, Россия, Москва, ул. Трубецкая, д. 8, стр. 2; <https://orcid.org/0000-0002-2990-7736>; VladzimirskijAV@zdrav.mos.ru

Омелянская Ольга Васильевна, руководитель по управлению подразделениями Дирекции Наука, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; <https://orcid.org/0000-0002-0245-4431>; OmelyanskayaOV@zdrav.mos.ru

Решетников Роман Владимирович, к.ф.-м.н., руководитель отдела научных медицинских исследований, Научно-практический клинический центр диагностики и телемедицинских технологий; 127051, Россия, Москва, ул. Петровка, д. 24, стр. 1; <https://orcid.org/0000-0002-9661-0254>; ReshetnikovRV1@zdrav.mos.ru

Information about the authors:

Yuri A. Vasilev, Dr. Sci. (Med.), Director, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; <https://orcid.org/0000-0002-5283-5961>; npcmr@zdrav.mos.ru

Ivan A. Blokhin, Cand. Sci. (Med.), Head of the Radiation Diagnostics Research Sector of the Department of Scientific Medical Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; <https://orcid.org/0000-0002-2681-9378>; BlokhinIA@zdrav.mos.ru

Maria R. Kodenko, Cand. Sci. (Eng.), Junior Researcher at the Department of Scientific Medical Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; Assistant Professor of the Department of Bio-medical Engineering Systems, Bauman Moscow State Technical University; 5, 2nd Baumanskaya St., Moscow, 105005, Russia; <https://orcid.org/0000-0002-0166-3768>; KodenkoMR@zdrav.mos.ru

Maria M. Suchilova, Junior Researcher at the Department of Scientific Medical Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; <https://orcid.org/0000-0003-1117-0294>; SuchilovaMM@zdrav.mos.ru

Anton V. Vladzimyrsky, Dr. Sci. (Med.), Deputy Director for Scientific Work, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; Professor of the Department of Information Technology and Medical Data Processing, Sechenov First Moscow State Medical University (Sechenov University); 8, Bldg. 2, Trubetskaya St., Moscow, 119991, Russia; <https://orcid.org/0000-0002-2990-7736>; VladzimirskijAV@zdrav.mos.ru

Olga V. Omelyanskaya, Head of Division Management of the Directorate of Science, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; <https://orcid.org/0000-0002-0245-4431>; OmelyanskayaOV@zdrav.mos.ru

Roman V. Reshetnikov, Cand. Sci. (Phys. & Math.), Head of the Department of Scientific Medical Research, Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies; 24, Bldg. 1, Petrovka St., Moscow, 127051, Russia; <https://orcid.org/0000-0002-9661-0254>; ReshetnikovRV1@zdrav.mos.ru