



Оригинальная статья

Согласованность между рентгенологами города Москвы при интерпретации маммографических исследований по шкале BI-RADS

Азарян А.С.^{1,2} • Пестренин Л.Д.¹ • Васильев Ю.А.¹ • Ахмад Е.С.¹ • Арзамасов К.М.¹

Азарян Авет Сергеевич – аспирант¹; врач-рентгенолог²;
ORCID: <https://orcid.org/0009-0007-8975-0017>
✉ 109457, г. Москва, ул. Федора Полетаева, 14/22–64, Российская Федерация.
E-mail: Dr.Azaryan@yandex.ru

Пестренин Лев Дмитриевич – мл. науч. сотр. отдела медицинской информатики, радиомики и радиогеномики¹;
ORCID: <http://orcid.org/0000-0002-1786-4329>.
E-mail: PestreninLD@zdrav.mos.ru

Васильев Юрий Александрович – канд. мед. наук, директор¹; ORCID: <http://orcid.org/0000-0002-5283-5961>. E-mail: npcmr@zdrav.mos.ru

Ахмад Екатерина Сергеевна – науч. сотр. сектора клинических и технических испытаний¹;
ORCID: <https://orcid.org/0000-0002-8235-9361>.
E-mail: AkhmadES@zdrav.mos.ru

Арзамасов Кирилл Михайлович – канд. мед. наук, руководитель отдела медицинской информатики, радиомики и радиогеномики¹;
ORCID: <http://orcid.org/0000-0001-7786-0349>.
E-mail: ArzamasovKM@zdrav.mos.ru

Актуальность. Злокачественные новообразования молочной железы занимают лидирующие позиции в структуре онкологической заболеваемости и смертности. Основным методом раннего выявления рака молочной железы признана маммография. Однако оценка маммографических исследований основана на субъективном мнении врача-рентгенолога, что может приводить к расхождению диагнозов. По данным литературы, согласованность рентгенологов при оценке маммограмм варьирует от 0,450 до 0,888.

Цель – оценить согласованность между врачами-рентгенологами г. Москвы при интерпретации маммографических исследований по шкале BI-RADS (Breast Imaging Reporting and Data System).

Материал и методы. Набор данных составил 741 маммографическое исследование, выполненное в период с 15.01.2020 по 25.06.2023. Все исследования были выгружены из Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы (ЕМИАС) г. Москвы (Россия) и содержали заключение врача-рентгенолога с оценкой по шкале BI-RADS (первичная оценка). Каждое исследование было дополнительно проанализировано двумя рентгенологами (с опытом работы от 2 до 5 лет) в рамках первого пересмотра и после еще двумя (с опытом работы более 5 лет и ученой степенью / званием) – в рамках экспертного пересмотра. Согласованность между врачами оценивали с помощью внутриклассового коэффициента корреляции.

Результаты. При оценке по полной шкале BI-RADS согласованность между врачами, выполнявшими первичную оценку, и врачами, выполнявшими первый пересмотр, находилась в диапазоне от 0,836 [95% доверительный

интервал (ДИ) 0,801–0,865] до 0,875 [95% ДИ 0,848–0,897]. Аналогичная согласованность наблюдалась между рентгенологами, выполнявшими первичную оценку, и экспертами: от 0,838 [95% ДИ 0,804–0,866] до 0,879 [95% ДИ 0,854–0,901]. При оценке по полной шкале BI-RADS согласованность между врачами, выполнявшими первый пересмотр, и экспертами оказалась статистически значимо выше ($p < 0,001$), чем с врачами, проводившими первичную оценку, и находилась в пределах от 0,890 [95% ДИ 0,866–0,910] до 0,963 [95% ДИ 0,954–0,970].

Заключение. Согласованность между врачами-рентгенологами в г. Москве при оценке маммографических исследований по полной шкале BI-RADS находится на высоком уровне. Согласованность между врачами-рентгенологами, выполнявшими пересмотр, выше, чем их согласованность с врачами, выполнявшими первичное описание, что может свидетельствовать о более качественных и стабильных результатах, получаемых в ходе пересмотра исследований.

Ключевые слова: рак молочной железы, маммография, скрининг, согласованность, BI-RADS

Для цитирования: Азарян АС, Пестренин ЛД, Васильев ЮА, Ахмад ЕС, Арзамасов КМ. Согласованность между рентгенологами города Москвы при интерпретации маммографических исследований по шкале BI-RADS. Альманах клинической медицины. 2024;52(7):377–384. doi: 10.18786/2072-0505-2024-52-035.

Поступила 03.09.2024; доработана 14.11.2024; принята к публикации 03.12.2024; опубликована онлайн 16.12.2024.

¹ ГБУЗ города Москвы «Научно-практический клинический центр диагностики и телемедицинских технологий Департамента здравоохранения города Москвы»; 127051, г. Москва, ул. Петровка, 24–1, Российская Федерация

² ГБУЗ города Москвы «Московский многопрофильный научно-клинический центр им. С.П. Боткина Департамента здравоохранения города Москвы»; 125284, г. Москва, 2-й Боткинский пр-д, 5, Российская Федерация

Злокачественные новообразования (ЗНО) молочной железы неизменно занимают второе место после рака легкого в структуре онкологической заболеваемости: так, в 2022 г. в мире было выявлено 2,3 млн новых случаев, что составило 11,6% от всех случаев впервые диагностированного рака. При этом среди женского населения

молочная железа по-прежнему остается ведущей локализацией ЗНО с долей 23,8%. В структуре смертности женщин от ЗНО наибольший удельный вес имеют также ЗНО молочной железы – 15,4% [1].

Основным скрининговым методом раннего выявления ЗНО молочной железы признана маммография [2]. Двойной просмотр результатов

профилактической маммографии, который принят на законодательном уровне в разных странах мира, включая Россию, повышает объединенную частоту обнаружения патологических изменений по сравнению с однократным просмотром [3]. При двойных просмотрах крайне необходимы стандартизация анализа и установление единой терминологии, которые обеспечиваются применением шкалы BI-RADS (Breast Imaging Reporting and Data System), разработанной Американским колледжем радиологии (American College of Radiology – ACR) [4, 5] и представленной в отечественных методических рекомендациях¹. Использование единой терминологии необходимо также для эффективной коммуникации между врачами разных специальностей. Кроме того, трудно переоценить значение стандартизированной терминологии как для сбора и анализа маммографических данных, так и для определения прогностической ценности находок в отношении злокачественности [6–9].

Несмотря на то что критерии установления категорий BI-RADS четко определены, применение этих критериев к маммографическим находкам основывается на их субъективной интерпретации врачом-рентгенологом. В связи с этим в ряде случаев может возникать несогласие между медицинскими специалистами при оценке одних и тех же маммографических изображений [10]. Маммограммы с искажением структуры молочной железы, кальцинатами, массовыми образованиями и участками асимметричного уплотнения чаще других вызывают несогласие между рентгенологами [11], при определении различных рентгенологических признаков коэффициент каппа Коэна варьирует от 0,090 до 0,820 [12]. В ряде исследований выявлено влияние пола врачей, рентгенологического стажа, оценки анамнеза исследуемого пациента и ряда других параметров на результат анализа маммографических изображений [13–17]. Как следствие, согласованность между рентгенологами при оценке маммограмм варьирует в широком диапазоне – от 0,450 до 0,888 в зависимости от типа используемой шкалы [18–22]. В случаях когда мнения рентгенологов расходятся, подразумевается, что как минимум один из двух рентгенологов, оценивающих маммографическое исследование, допускает ошибку. По мнению А.Р. Brady, полностью исключить ошибки

в интерпретации рентгенологических исследований невозможно, однако реально снизить частоту их возникновения, что могло бы способствовать повышению согласованности между рентгенологами [23]. В целом оценка согласованности необходима для того, чтобы можно было сформировать четкое представление о возможных вариациях при интерпретации того или иного вида рентгенологического исследования [24].

Таким образом, зарубежными исследователями была выявлена проблема отсутствия высокой согласованности между врачами-рентгенологами при интерпретации маммографических исследований, однако в отечественной литературе нам не удалось найти научные работы, посвященные оценке согласованности рентгенологов, которые были бы выполнены на достаточных по объему выборках.

Цель настоящего исследования – оценить согласованность между врачами-рентгенологами г. Москвы при интерпретации маммографических исследований по шкале BI-RADS.

Материал и методы

Набор данных

Из Единого радиологического информационного сервиса Единой медицинской информационно-аналитической системы (ЕРИС ЕМИАС) г. Москвы были выгружены маммографические исследования, выполненные в период с 15.01.2020 по 25.06.2023, которые содержали заключение врача-рентгенолога с указанием категории для правой и левой молочной железы по шкале BI-RADS (первичная оценка).

Критерии включения: исследование, выполненное пациентам женского пола в возрасте старше 18 лет, наличие выставленной при первичной оценке категории BI-RADS.

Критерии исключения: маммограммы неудовлетворительного качества (класс I по PGMI (Perfect, Good, Moderate, Inadequate)).

В итоговый набор данных вошло 741 исследование. Средний возраст пациенток составил $59,5 \pm 12,4$ года. Процентное распределение исследований по категориям BI-RADS по результатам первичной оценки для правой и левой молочной железы статистически значимо не отличалось ($p > 0,05$) (табл. 1).

Таблица 1. Распределение исследований по категориям BI-RADS по результатам первичной оценки, абс. (%)

Локализация	BI-RADS 1	BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5
Правая молочная железа	223 (30,1)	306 (41,3)	97 (13,1)	50 (6,8)	64 (8,7)
Левая молочная железа	199 (26,9)	330 (44,5)	76 (10,2)	54 (7,3)	82 (11,1)



Оценка маммографических исследований

Каждое исследование оценивали четыре рентгенолога.

На первом этапе каждое исследование было проанализировано независимо друг от друга двумя врачами-рентгенологами с опытом работы от 2 до 5 лет. В процессе оценки исследований каждый рентгенолог определял категорию по шкале BI-RADS для правой и левой молочной железы и вносил данные в таблицу (первый пересмотр). Оценку категории по шкале BI-RADS выполняли на основании характерных рентгенологических признаков^{2,3}.

На втором этапе другие два врача-рентгенолога с опытом работы не менее 5 лет и ученой степенью / званием также независимо друг от друга определяли категорию по шкале BI-RADS для правой и левой молочной железы (экспертный пересмотр). На этом этапе врачам-рентгенологам были доступны результаты оценки, внесенные в таблицу на первом этапе.

Статистический анализ

Статистический анализ выполнен в интегрированной среде разработки RStudio с использованием языка R. Для оценки согласованности между врачами-рентгенологами использовали метод оценки внутриклассовой корреляции Пирсона [25, 26]. Согласованность рассчитывали для полной шкалы BI-RADS (категории 1–5), а также для двух бинарных шкал. Бинарная шкала 1 включала в себя категории «отсутствие признаков ЗНО» (категории BI-RADS 1, 2, 3) и «признаки ЗНО» (категории BI-RADS 4 и 5) [27]. Бинарная шкала 2 включала в себя категории «отсутствие признаков ЗНО» (категории BI-RADS 1 и 2) и «признаки ЗНО» (категории BI-RADS 3, 4, 5). Размер

выборки обусловлен имеющимися в наличии исследованиями, полученными при подготовке набора данных. На основании метода определения объема выборки, предложенного D.G. Bonett, для обеспечения уровня значимости, равного 0,05, при определении согласованности двух рентгенологов минимальный объем выборки должен составлять 377 исследований [28].

Результаты

Согласованность между врачами, выполнявшими первичную оценку, и врачами, проводившими первый пересмотр, находилась в пределах 0,836–0,875 при оценке по полной шкале BI-RADS. Аналогичная согласованность наблюдалась между врачами, выполнявшими первичную оценку, и экспертами (0,838–0,879) (табл. 2).

Согласованность по бинарной шкале 1 для всех сравниваемых оценок была сопоставимой с согласованностью по полной шкале BI-RADS или в ряде случаев превышала ее. Согласованность по бинарной шкале 2 для всех сравниваемых оценок была сопоставимой с согласованностью по полной шкале BI-RADS или, наоборот, в ряде случаев была ниже ее.

Мы оценили согласованность между врачами и экспертами, участвовавшими в пересмотре исследований. Коэффициент внутриклассовой корреляции был в пределах от 0,890 до 0,963 (табл. 3).

Согласованность между врачами, выполнявшими первый пересмотр, и экспертами по полной шкале BI-RADS в большинстве случаев была значимо выше, чем согласованность их оценок и оценок, полученных в процессе первичного описания маммографических исследований.

Таблица 2. Согласованность между врачами-рентгенологами, выполнявшими первичное описание, первый и экспертный пересмотры

Сравниваемые оценки		Молочная железа	Согласованность по полной шкале BI-RADS	Согласованность по бинарной шкале 1	Согласованность по бинарной шкале 2
Первичная оценка	Врач 1 (первый пересмотр)	R	0,836 [0,801–0,865]	0,831 [0,796–0,861]	0,777 [0,732–0,815]
		L	0,875 [0,848–0,897]	0,923 [0,906–0,937]	0,785 [0,741–0,822]
	Врач 2 (первый пересмотр)	R	0,847 [0,815–0,874]	0,810 [0,771–0,843]	0,786 [0,743–0,823]
		L	0,873 [0,846–0,896]	0,941 [0,928–0,952]	0,734 [0,683–0,779]
	Эксперт 1	R	0,847 [0,814–0,874]	0,831 [0,796–0,861]	0,792 [0,750–0,828]
		L	0,879 [0,854–0,901]	0,923 [0,906–0,937]	0,770 [0,725–0,810]
	Эксперт 2	R	0,838 [0,804–0,866]	0,844 [0,811–0,871]	0,770 [0,725–0,810]
		L	0,862 [0,832–0,886]	0,923 [0,906–0,937]	0,829 [0,801–0,855]

L – левая молочная железа, R – правая молочная железа

Все результаты представлены в виде коэффициентов внутриклассовой корреляции с 95% доверительным интервалом; значение p для всех сравниваемых оценок составило $< 0,001$

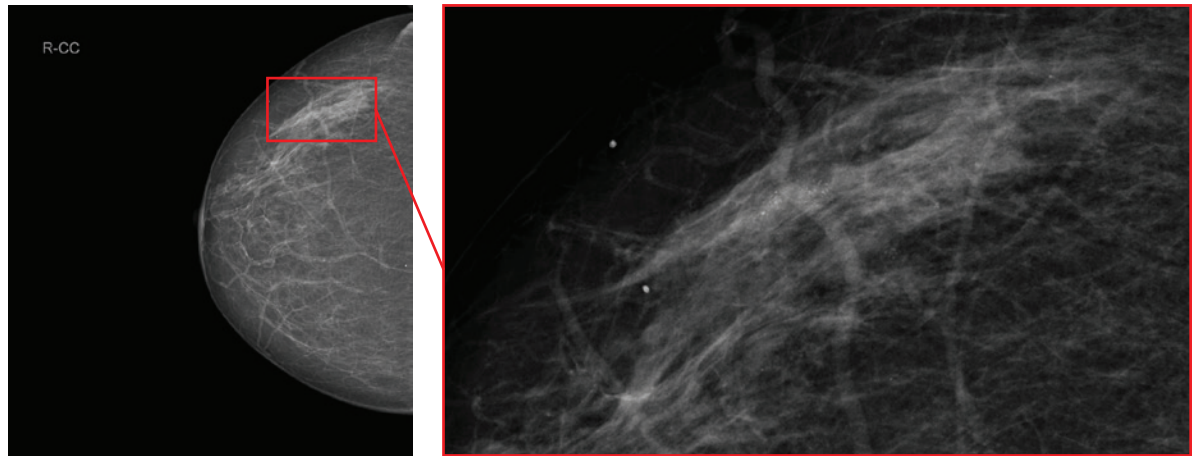


Рис. 1. Правая молочная железа в краниокаудальной проекции с выделенной областью интереса. Очаговое скопление микрокальцинатов в паренхиме правой молочной железы

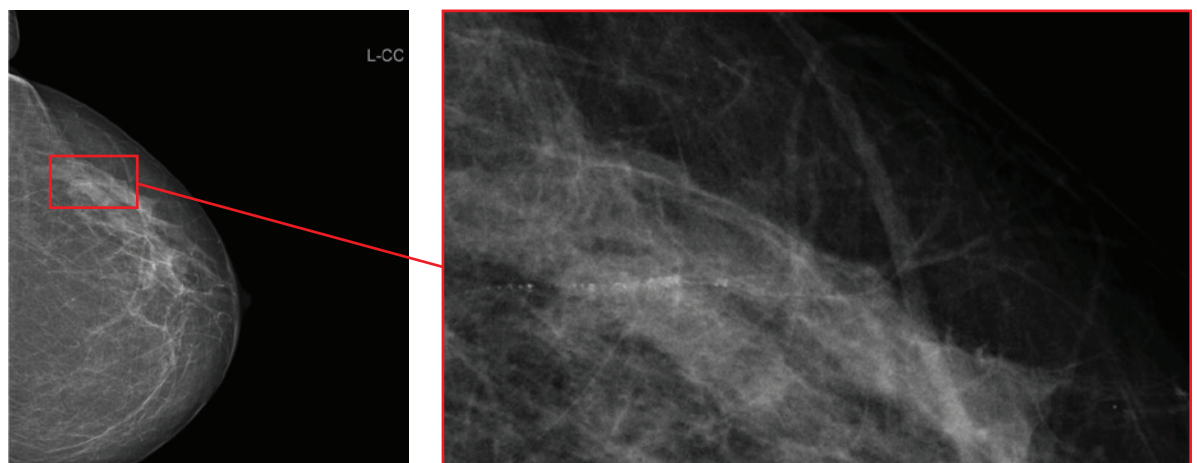


Рис. 2. Левая молочная железа в краниокаудальной проекции с выделенной областью интереса. Мнения рентгенологов разделились: в области интереса – очаговое скопление микрокальцинатов или линейная фрагментарная кальцинация в проекции стенки сосуда

Однако присутствовали и единичные случаи значимого расхождения в оценках маммографических исследований. В одном из таких случаев рентгенолог, выполнивший первичное описание, и врачи, выполнившие пересмотр, обратили внимание на очаговое скопление микрокальцинатов в паренхиме правой молочной железы (рис. 1). Все врачи оценили изменения как соответствовавшие категории BI-RADS 4. Не исключено, что выявленная патология в правой молочной железе повлияла на оценку левой молочной железы у той же пациентки, где двое рентгенологов распознали в указанной на рис. 2 зоне интереса очаговое скопление микрокальцинатов и определили категорию BI-RADS 4, тогда как остальные врачи интерпретировали изменения как линейную фрагментарную кальцинацию в проекции стенки сосуда и установили категорию BI-RADS 2.

Обсуждение

По нашим данным, согласованность при оценке по полной шкале BI-RADS находилась в диапазоне от 0,836 до 0,963, а при оценке по бинарным шкалам – от 0,734 до 0,941.

В исследованиях, выполненных другими авторами, согласованность варьировала в пределах от 0,450 до 0,888 (табл. 4) [20–24]. Обращают на себя внимание три фактора, которые не позволяют в полной мере выполнить сопоставление полученных в этих исследованиях результатов, а также сравнить их с нашими данными. Во-первых, выбирались разные шкалы оценки маммографических исследований: шкалы BI-RADS, включавшие 4, 5 или 6 категорий; шкалы злокачественности, включавшие 2 и 7 категорий; шкала «патология / без патологии»; шкала «пересмотр

**Таблица 3.** Согласованность между врачами-рентгенологами, выполнявшими первый и экспертный пересмотры

Сравниваемые оценки		Молочная железа	Согласованность по полной шкале BI-RADS	Согласованность по бинарной шкале 2
Врач 1 (первый пересмотр)	Эксперт 1	R	0,963 [0,954–0,970]	0,932 [0,917–0,944]
		L	0,954 [0,944–0,963]	0,934 [0,919–0,946]
	Эксперт 2	R	0,904 [0,882–0,921]	0,862 [0,833–0,887]
		L	0,890 [0,866–0,910]	0,934 [0,919–0,946]
Врач 2 (первый пересмотр)	Эксперт 1	R	0,924 [0,907–0,938]	0,910 [0,890–0,926]
		L	0,929 [0,913–0,942]	0,886 [0,862–0,907]
	Эксперт 2	R	0,958 [0,948–0,966]	0,971 [0,965–0,977]
		L	0,958 [0,949–0,966]	0,886 [0,862–0,907]
Врач 1 (первый пересмотр)	Врач 2 (первый пересмотр)	R	0,918 [0,900–0,933]	0,891 [0,868–0,911]
		L	0,903 [0,882–0,921]	0,824 [0,788–0,855]
Эксперт 1	Эксперт 2	R	0,910 [0,890–0,926]	0,880 [0,854–0,902]
		L	0,914 [0,896–0,930]	0,886 [0,860–0,906]

L – левая молочная железа, R – правая молочная железа

Все результаты представлены в виде коэффициентов внутриклассовой корреляции с 95% доверительным интервалом; значение *p* для всех сравниваемых оценок составило < 0,001

(BI-RADS 0, 4 и 5) / без пересмотра (BI-RADS 1 или 2)» [18–22]. Во-вторых, результаты исследований опубликованы в широком временном диапазоне – от 1998 до 2019 гг., что, по нашему мнению, имеет принципиальное значение, поскольку внедрение скрининговой маммографии и обучение врачей-рентгенологов правильной интерпретации таких исследований происходили в разное время в зависимости от страны, что не могло не сказаться на точности и согласованности врачебных оценок [29]. В-третьих, со временем менялась и сама шкала BI-RADS. Всего было выпущено 5 изданий: в 1993, 1995, 1998, 2003 и 2013 гг. [30]. В ряде исследований отсутствовали указания на то, каким изданием шкалы BI-RADS пользовались врачи-рентгенологи

при оценке маммографических исследований [18, 20, 21].

Тем не менее при анализе работ других авторов мы обнаружили интересную закономерность: согласованность между врачами-рентгенологами была выше при оценке по бинарной шкале, чем при оценке по полной шкале BI-RADS [18, 20, 22]. В нашем исследовании мы также установили, что согласованность оценок по бинарной шкале 1 «отсутствие признаков ЗНО (BI-RADS 1, 2, 3) / признаки ЗНО (BI-RADS 4 и 5)» была несколько выше, чем согласованность по полной шкале BI-RADS, включавшей в себя 5 категорий (0,810–0,941 против 0,836–0,879 соответственно) (см. табл. 1). Однако согласованность по бинарной

Таблица 4. Согласованность между врачами-рентгенологами по данным литературы

Страна	Год публикации	Тип шкалы	Согласованность	95% доверительный интервал	Источник
Республика Корея	2019	Шкала злокачественности (7 категорий)	0,772	0,754–0,791	[18]
		Шкала злокачественности (2 категории)	0,888	0,873–0,902	
Дания	2015	BI-RADS (4 категории)	0,880	0,810–0,920	[19]
Голландия	2013	BI-RADS (5 категорий)	0,450	0,430–0,470	[20]
		Бинарная шкала (BI-RADS 0, 4 и 5 / BI-RADS 1 и 2)	0,770	0,760–0,790	
Испания	2012	BI-RADS (6 категорий)	0,580	0,560–0,590	[21]
США	1998	BI-RADS (5 категорий)	0,580	0,550–0,610	[22]
		Бинарная шкала (патология / без патологии)	0,660	0,630–0,690	



шкале 2 «отсутствие признаков ЗНО (BI-RADS 1 и 2) / признаки ЗНО (BI-RADS 3, 4, 5)», наоборот, была несколько ниже, чем согласованность по полной шкале BI-RADS, включавшей в себя 5 категорий (0,734–0,829 против 0,836–0,879 соответственно), а также ниже, чем согласованность по бинарной шкале 1 (0,734–0,829 против 0,810–0,941 соответственно). В этих случаях разную согласованность можно объяснить отнесением категории BI-RADS 3 к категории с наличием или отсутствием признаков ЗНО. Эта категория часто устанавливается при наличии изменений и отсутствии архива предыдущих исследований у пациента, но тем не менее согласно шкале BI-RADS выявленные изменения относятся скорее к доброкачественным, нежели злокачественным находкам. Вероятно, вследствие этого согласованность по бинарной шкале 1 оказалась выше, чем по бинарной шкале 2.

Ограничением настоящего исследования было отсутствие гистологической верификации заключений врачей-рентгенологов.

Заключение

Согласованность между врачами-рентгенологами в г. Москве при оценке маммографических исследований по полной шкале BI-RADS находится на высоком уровне и составляет от 0,836 до 0,963. Согласованность между врачами-рентгенологами, выполнявшими пересмотр (0,890–0,963), выше, чем их согласованность с врачами, выполнявшими первичное описание (0,836–0,879), что может свидетельствовать о более качественных и стабильных результатах, получаемых в ходе пересмотра исследований.

Оценка согласованности врачей-рентгенологов по-прежнему представляет значительный интерес для исследователей всего мира. Логичным продолжением настоящей работы могут стать более крупные исследования, посвященные оценке согласованности врачей-рентгенологов разных стран при установлении категории BI-RADS, которые необходимо провести на гистологически верифицированных наборах данных. ©

¹ Мануйлова ОО, Павлова ТВ, Диденко ВВ, Смирнов ИВ, Абдураимов АБ, Васильев АЮ. Методические рекомендации по использованию международной системы BI-RADS при маммографическом обследовании. Москва, 2017. 23 с.

² Терновой СК, Абдураимов АБ. Лучевая маммология. М.: ГЭОТАР-Медиа; 2007.

³ Морозов СП, Ветшева НН, Диденко ВВ, Смирнов ИВ, Овсянников АГ, Ледихова НВ, Пучкова ОС. Организация программы популяционного скрининга злокачественных новообразований молочной железы среди женского населения: методические рекомендации. Москва, 2020. 44 с.

Дополнительная информация

Финансирование

Данная статья подготовлена авторским коллективом в рамках НИОКР «Разработка платформы подготовки наборов данных лучевых диагностических исследований» (№ ЕГИСУ: 123031500003-8) в соответствии с Приказом от 21.12.2022 г. № 1196 «Об утверждении государственных заданий, финансовое обеспечение которых осуществляется за счет средств бюджета города Москвы, государственным бюджетным (автономным) учреждениям, подведомственным Департаменту здравоохранения города Москвы, на 2023 год и плановый период 2024 и 2025 годов» Департамента здравоохранения города Москвы.

Конфликт интересов

Авторы заявляют об отсутствии явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Участие авторов

А.С. Азарян – анализ полученных данных, написание и редактирование текста; Л.Д. Пестренин – сбор и обработка материала, анализ полученных данных, написание и финальное редактирование текста; Ю.А. Васильев – концепция исследования, финальное редактирование текста; Е.С. Ахмад, К.М. Арзамасов – дизайн исследования, сбор и обработка материала. Все авторы прочли

и одобрили финальную версию статьи перед публикацией, согласны нести ответственность за все аспекты работы и гарантируют, что ими надлежащим образом были рассмотрены и решены вопросы, связанные с точностью и добросовестностью всех частей работы.

Благодарности

Авторы выражают благодарность А.В. Владимировскому и О.В. Омелянской за помощь в создании концепции исследования, Н.Ю. Никитину и А.В. Майоровой – за консультирование по статистической обработке данных, В.В. Зинченко – за помощь с организацией сбора данных, С.С. Семенову – за подготовку первичных данных, М.Ю.Болисовой – за консультации по клиническим вопросам.

Список литературы / References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, Jemal A. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74(3):229–263. doi: 10.3322/caac.21834.
2. Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: A systematic review. *Breast.* 2022;64:85–99. doi: 10.1016/j.breast.2022.04.003.
3. Васильев ЮА, Тыров ИА, Владимировский АВ, Арзамасов КМ, Шулькин ИМ, Кожихина ДД, Пестренин ЛД. Двойной просмотр результатов маммографии с применением технологий искусственного интеллекта: новая модель организации массовых профилактических исследований. *Digital Diagnostics.* 2023;4(2):93–104. doi: 10.17816/DD321423. Vasilev YA, Tyrov IA, Vladzimirskyy AV, Arzamasov KM, Shulkin IM, Kozhikhina DD, Pestrenin LD. [Double-reading mammograms using artificial intelligence technologies: A new model of mass preventive examination organization]. *Digital Diagnostics.*



- 2023;4(2):93–104. Russian. doi: 10.17816/DD321423.
4. D'Orsi CJ. The American College of Radiology mammography lexicon: An initial attempt to standardize terminology. *AJR Am J Roentgenol.* 1996;166(4):779–780. doi: 10.2214/ajr.166.4.8610548.
 5. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). *Radiol Clin North Am.* 2002;40(3):409–430, v. doi: 10.1016/s0033-8389(01)00017-3.
 6. Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith-Bindman R. Positive predictive value of specific mammographic findings according to reader and patient variables. *Radiology.* 2009;250(3):648–657. doi: 10.1148/radiol.2503080541.
 7. Liberman L, Abramson AF, Squires FB, Glassman JR, Morris EA, Dershaw DD. The breast imaging reporting and data system: Positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol.* 1998;171(1):35–40. doi: 10.2214/ajr.171.1.9648759.
 8. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. *AJR Am J Roentgenol.* 2010;194(5):1378–1383. doi: 10.2214/AJR.09.3423.
 9. Leung JW, Sickles EA. Developing asymmetry identified on mammography: Correlation with imaging outcome and pathologic findings. *AJR Am J Roentgenol.* 2007;188(3):667–675. doi: 10.2214/AJR.06.0413.
 10. Васильев ЮА, Владимирский АВ, Арзамасов КМ, Шулькин ИМ, Аксенова ЛЕ, Пестренин ЛД, Семенов СС, Бондарчук ДВ, Смирнов ИВ. Первые 10 000 маммографических исследований, выполненных в рамках услуги «Описание и интерпретация данных маммографического исследования с использованием искусственного интеллекта». *Менеджер здравоохранения.* 2023;(8):54–67. doi: 10.21045/1811-0185-2023-8-54-67.
Vasiliev YA, Vladzimirsky AV, Arzamasov KM, Shulkin IM, Aksenova LE, Pestrenin LD, Semenov SS, Bondarchuk DV, Smirnov IV. [The first 10,000 mammography exams performed as part of the "Description and interpretation of mammography data using artificial intelligence" service]. *Manager Zdravookhranenia.* 2023;(8):54–67. Russian. doi: 10.21045/1811-0185-2023-8-54-67.
 11. Beam CA, Conant EF, Sickles EA. Factors affecting radiologist inconsistency in screening mammography. *Acad Radiol.* 2002;9(5):531–540. doi: 10.1016/s1076-6332(03)80330-6.
 12. Lee AY, Wisner DJ, Aminololama-Shakeri S, Arasu VA, Feig SA, Hargreaves J, Ojeda-Fournier H, Bassett LW, Wells CJ, De Guzman J, Flowers CI, Campbell JE, Elson SL, Retallack H, Joe BN. Inter-reader variability in the use of BI-RADS descriptors for suspicious findings on diagnostic mammography: A multi-institution study of 10 academic radiologists. *Acad Radiol.* 2017;24(1):60–66. doi: 10.1016/j.acra.2016.09.010.
 13. Barlow WE, Chi C, Carney PA, Taplin SH, D'Orsi C, Cutter G, Hendrick RE, Elmore JG. Accuracy of screening mammography interpretation by characteristics of radiologists. *J Natl Cancer Inst.* 2004;96(24):1840–1850. doi: 10.1093/jnci/djh333.
 14. Carney PA, Elmore JG, Abraham LA, Gerrity MS, Hendrick RE, Taplin SH, Barlow WE, Cutter GR, Poplack SP, D'Orsi CJ. Radiologist uncertainty and the interpretation of screening. *Med Decis Making.* 2004;24(3):255–264. doi: 10.1177/0272989X04265480.
 15. Elmore JG, Jackson SL, Abraham L, Miglioretti DL, Carney PA, Geller BM, Yankaskas BC, Kerlikowske K, Onega T, Rosenberg RD, Sickles EA, Buist DS. Variability in interpretive performance at screening mammography and radiologists' characteristics associated with accuracy. *Radiology.* 2009;253(3):641–651. doi: 10.1148/radiol.2533082308.
 16. Reed WM, Lee WB, Cawson JN, Brennan PC. Malignancy detection in digital mammograms: Important reader characteristics and required case numbers. *Acad Radiol.* 2010;17(11):1409–1413. doi: 10.1016/j.acra.2010.06.016.
 17. Carney PA, Cook AJ, Miglioretti DL, Feig SA, Bowles EA, Geller BM, Kerlikowske K, Kettler M, Onega T, Elmore JG. Use of clinical history affects accuracy of interpretive performance of screening mammography. *J Clin Epidemiol.* 2012;65(2):219–230. doi: 10.1016/j.jclinepi.2011.06.010.
 18. Kim SH, Lee EH, Jun JK, Kim YM, Chang YW, Lee JH, Kim HW, Choi EJ; Alliance for Breast Cancer Screening in Korea (ABCS-K). Interpretive performance and inter-observer agreement on digital mammography test sets. *Korean J Radiol.* 2019;20(2):218–224. doi: 10.3348/kjr.2018.0193.
 19. Winkel RR, von Euler-Chelpin M, Nielsen M, Diao P, Nielsen MB, Uldall WY, Vejborg I. Inter-observer agreement according to three methods of evaluating mammographic density and parenchymal pattern in a case control study: Impact on relative risk of breast cancer. *BMC Cancer.* 2015;15:274. doi: 10.1186/s12885-015-1256-3.
 20. Timmers JM, Verbeek AL, Pijnappel RM, Broeders MJ, den Heeten GJ. Experiences with a self-test for Dutch breast screening radiologists: Lessons learnt. *Eur Radiol.* 2014;24(2):294–304. doi: 10.1007/s00330-013-3018-4.
 21. Redondo A, Comas M, Macià F, Ferrer F, Murta-Nascimento C, Maristany MT, Molins E, Sala M, Castells X. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *Br J Radiol.* 2012;85(1019):1465–1470. doi: 10.1259/bjr/21256379.
 22. Kerlikowske K, Grady D, Barclay J, Frankel SD, Ominsky SH, Sickles EA, Ernster V. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. *J Natl Cancer Inst.* 1998;90(23):1801–1809. doi: 10.1093/jnci/90.23.1801.
 23. Brady AP. Error and discrepancy in radiology: Inevitable or avoidable? *Insights Imaging.* 2017;8(1):171–182. doi: 10.1007/s13244-016-0534-1-1.
 24. Benchoufi M, Matzner-Lober E, Molinari N, Jannot AS, Soyer P. Interobserver agreement issues in radiology. *Diagn Interv Imaging.* 2020;101(10):639–641. doi: 10.1016/j.diii.2020.09.001.
 25. Müller R, Büttner P. A critical discussion of intraclass correlation coefficients. *Stat Med.* 1994;13(23-24):2465–2476. doi: 10.1002/sim.4780132310.
 26. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1(1):30–46. doi: 10.1037/1082-989X.1.1.30.
 27. Taplin SH, Ichikawa LE, Kerlikowske K, Ernster VL, Rosenberg RD, Yankaskas BC, Carney PA, Geller BM, Urban N, Dignan MB, Barlow WE, Ballard-Barbash R, Sickles EA. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology.* 2002;222(2):529–535. doi: 10.1148/radiol.2222010647.
 28. Bonett DG. Sample size requirements for estimating intraclass correlations with desired precision. *Stat Med.* 2002;21(9):1331–1335. doi: 10.1002/sim.1108.
 29. Trieu PDY, Mello-Thoms CR, Barron ML, Lewis SJ. Look how far we have come: BREAST cancer detection education on the international stage. *Front Oncol.* 2023;12:1023714. doi: 10.3389/fonc.2022.1023714.
 30. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, Mendelson EB, Wilcox PA, Butler PF, D'Orsi CJ. The ACR BI-RADS experience: learning from history. *J Am Coll Radiol.* 2009;6(12):851–860. doi: 10.1016/j.jacr.2009.07.023.



The inter-reader agreement in the interpretation of mammography images according to BI-RADS by Moscow radiologists

A.S. Azaryan^{1,2} • L.D. Pestrenin¹ • Yu.A. Vasilev¹ • E.S. Ahmad¹ • K.M. Arzamasov¹

Background: Breast malignancies take a leading position among incident cancers in women. Mammography has been recognized as the main method for early detection of breast cancer. However, mammogram assessments are based on a subjective opinion of the radiologist, which could lead to diagnostic disagreement. According to the literature, inter-radiologist agreement on mammograms varies from 0.450 to 0.888.

Aim: To assess the inter-reader agreement in mammogram interpretation with BI-RADS (Breast Imaging Reporting and Data System) by radiologists of the Moscow city (Russia).

Methods: The study included 741 mammography images done from January 15, 2020, to June 25, 2023. All mammograms were downloaded from the Unified Radiology Information Service of the Unified Medical Information and Analytical System (EMIAS) of the Moscow city and included radiologist reports with a BI-RADS score (the initial assessment). Each mammogram was further analyzed by two radiologists (with their job experience from 2 to 5 years) (this was the first revision) and thereafter by two more radiologists (with their job experience above 5 years and scientific degree) as a part of the expert review. The inter-reader agreement was assessed using an intra-class correlation coefficient.

Results: The inter-reader agreement for the full BI-RADS score between radiologists who performed the initial assessment and those performing the first revision ranged from 0.836 [95% confidence interval (CI) 0.801–0.865] to 0.875 [95%

CI 0.848–0.897]. Similar agreement was observed between radiologists who performed the initial assessment and the experts: 0.838 [95% CI 0.804–0.866] to 0.879 [95% CI 0.854–0.901]. The agreement on the full BI-RADS scale between radiologists who performed the first revision and the experts was significantly higher ($p < 0.001$) than with those performing the initial assessment and ranged from 0.890 [95% CI 0.866–0.910] to 0.963 [95% CI 0.954–0.970].

Conclusion: The inter-reader agreement between radiologists of the Moscow city in the assessment of mammography study results on the full BI-RADS scale is high. The agreement between the radiologists who performed the revision is higher than their agreement with the radiologists who performed the initial assessment, which may indicate better and more stable results obtained during the revision.

Key words: breast cancer, mammography, screening, agreement, BI-RADS

For citation: Azaryan AS, Pestrenin LD, Vasilev YuA, Ahmad ES, Arzamasov KM. The inter-reader agreement in the interpretation of mammography images according to BI-RADS by Moscow radiologists. *Almanac of Clinical Medicine*. 2024;52(7):377–384. doi: 10.18786/2072-0505-2024-52-035.

Received 3 September 2024; revised 14 November 2024; accepted 3 December 2024; published online 16 December 2024

Funding

This paper was prepared by a group of authors as part of the research and development effort titled "Development of a platform to generate data sets containing diagnostic imaging studies" (USIS No.: 123031500003-8) in accordance with the Order No. 1196 dated December 21, 2022 "On approval of state assignments funded by means of allocations from the budget of the city of Moscow to the state budgetary (autonomous) institutions subordinate to the Moscow Health Care Department, for 2023 and the planned period of 2024 and 2025" issued by the Moscow Health Care Department.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this article.

Authors' contribution

A.S. Azaryan, data analysis, text writing and editing; L.D. Pestrenin, data collection and management, data analysis, text writing and final editing; Yu.A. Vasilev, the study concept, final text editing; E.S. Ahmad, K.M. Arzamasov, the study design, data collection and management. All the authors have read and approved the final version of the manuscript before submission, agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgement

The authors would like to acknowledge A.V. Vladzimirskyy and O.V. Omelyanskaya for their assistance with the study concept, N.Yu. Nikitin and A.V. Maiorova for their statistical advice, V.V. Zinchenko for the assistance with the data collection, S.S. Semenov for preparation of the primary data, M.Yu. Bolisova for her clinical advice.

Avet S. Azaryan – Postgraduate Student¹; Radiologist²; ORCID: <https://orcid.org/0009-0007-8975-0017>

✉ Ul. Fedora Poletaeva 14/22–64, Moscow, 109457, Russian Federation. E-mail: Dr.Azaryan@yandex.ru

Lev D. Pestrenin – Junior Research Fellow, Department of Medical Informatics, Radiomics and Radiogenomics¹; ORCID: <http://orcid.org/0000-0002-1786-4329>. E-mail: PestreninLD@zdrav.mos.ru

Yuriy A. Vasilev – MD, PhD, Director¹; ORCID: <http://orcid.org/0000-0002-5283-5961>. E-mail: npcmr@zdrav.mos.ru

Ekaterina S. Akhmad – Research Fellow, Clinical and Technical Trials Sector¹; ORCID: <https://orcid.org/0000-0002-8235-9361>. E-mail: AkhmadES@zdrav.mos.ru

Kirill M. Arzamasov – MD, PhD, Head of the Department of Medical Informatics, Radiomics and Radiogenomics¹; ORCID: <http://orcid.org/0000-0001-7786-0349>. E-mail: ArzamasovKM@zdrav.mos.ru

¹Research and Practical Clinical Center for Diagnostics and Telemedicine Technologies of Moscow Health Care Department; ul. Petrovka 24–1, Moscow, 127051, Russian Federation

²Botkin Hospital; 2nd Botkinsky pr-d 5, Moscow, 125284, Russian Federation