Annals of the Russian Academy of Medical Sciences. 2022;77(2):97–106.

# С.С. Первушкин, П.М. Зельтер, Е.К. Крамм, Е.А. Сартакова



Самарский государственный медицинский университет, Самара, Российская Федерация

# Сравнение уровня экспертного согласия в оценке объема поражения при COVID-19-ассоциированной пневмонии на компьютерной томографии грудной клетки

Обоснование. Пандемия SARS-CoV-2 поставила перед специалистами по лучевой диагностике новые задачи — ранней диагностики интерстициальной пневмонии, оценки ее тяжести и динамики при контрольных исследованиях. Компьютерная томография (КТ) стала методом выбора для оценки легочной ткани у больных COVID-19, что связано с низкой чувствительностью рентгенографии в детекции снижения воздушности по типу «матового стекла». Критерии оценки визуальных признаков поражения легочной ткани зачастую обладают долей субъективности, при этом заключение, основанное на них, может влиять на тактику ведения пациента. **Цель исследования** — определить меж- и внутриэкспертную сходимость в оценке процента поражения и KT-степень COVID-19-ассоциированной пневмонии при КТ-исследовании среди экспертов с разным опытом, проанализировать уровень сходимости в зависимости от распространенности поражения и иных факторов. Методы. В исследовании проанализированы КТ 221 пациента с подтвержденным ПЦР к SARS-CoV-2. Исключены пациенты с фоновой патологией легких и часть пациентов с поражением до 50% для создания равномерной по степеням выборке из 100 пациентов. Четыре эксперта-рентгенолога определяли процент поражения легких и КТ-степень. Произведен анализ результатов экспертной оценки методами классической описательной статистики и анализ внутри- и межэкспертной сходимости. **Результаты.** При оценке внутриэкспертной сходимости (через 6 мес) корреляция уровня поражения легких в процентах между первым и вторым чтением составила у эксперта 1 (высокий уровень подготовки)  $\mathit{R} = \mathit{0,86}$  (р  $< \mathit{0,05}$ ), у эксперта  $\mathit{2}$ (высокий уровень подготовки) — R=0.84 (р < 0.05). Внутриэкспертное согласие Kanna (для KT-степени) составило 0.54 для эксперта 1 и 0,46 для эксперта 2, что соответствует умеренному уровню согласованности. При оценке межэкспертной сходимости взаимосвязь между уровнем поражения легких в процентах между экспертами 1 и 2 (высокий уровень подготовки) составила R=0.87 (p<0.05), между экспертами 3 и 4 (низкий уровень подготовки) — R = 0.78 (р < 0.05). Мера межэкспертного согласия Каппа составила 0.51 для экспертов 1 и 2 и 0,56 — для экспертов 3 и 4. Средняя оценка экспертов по выборке различалась до 4,5%, а при анализе расхождений мнений экспертов разница варьировалась равномерно, в степень и как увеличения, так и уменьшения объема поражения, тем самым представляя собой случайную ошибку. Заключение. Несмотря на высокий уровень корреляции в оценке процента поражения, сходимость Каппа по КТ-степени оказалась умеренной, существенно не различаясь от степени подготовки эксперта. Чаще различия в уровне поражения наблюдаются при «неклассических» паттернах: «обратном гало», криволинейных уплотнениях и т.д. Разница в мнениях не представляет собой систематическую ошибку. Таким образом, экспертная оценка объема поражения легких «эмпирически» обладает умеренным, недостаточным уровнем надежности, что, в свою очередь, создает благоприятные условия для использования инструментов по объективизации оценки.

**Ключевые слова**: COVID-19, пневмония, внутриэкспертная надежность, межэкспертная надежность

Для цитирования: Первушкин С.С., Зельтер П.М., Крамм Е.К., Сартакова Е.А. Сравнение уровня экспертного согласия в оценке объема поражения при COVID-19-ассоциированной пневмонии на компьютерной томографии грудной клетки. Вестник РАМН. 2022;77(2):97-106. doi: https://doi.org/10.15690/vramn1644

# Введение

Пандемия SARS-CoV-2 поставила перед врачамиспециалистами новые задачи, такие как ранняя диагностика интерстициальной пневмонии, оценка ее тяжести и динамики при контрольных исследованиях [1]. Компьютерная томография (КТ) стала методом выбора для оценки легочной ткани у больных COVID-19, что связано с низкой чувствительностью рентгенографии в детекции снижения воздушности по типу «матового стекла» [2]. Изменения легочной ткани, выявляемые при КТ, подробно описаны и представляют собой снижение воздушности по типу «матового стекла» и консолидации и другие менее частые паттерны [3]. В разных странах приняты различные принципы описания КТ при подозрении на интерстициальную пневмонию. Внедряются в практику различные системы по оценке вероятности COVID19-пневмонии: COVID-RADS и CoRADS [4, 5].

Некоторые исследователи полагают верным называть поражение легких при COVID-19 не пневмонией, а пульмонитом. Такой подход базируется на том, что основным патологоанатомическим проявлением в легких является диффузное альвеолярное повреждение в сочетании с альвеолярно-геморрагическим синдромом [15]. При этом в лучевой диагностике общепринято использование термина «пневмония» в связи с тем, что в целом томографические паттерны соответствуют картине при других пневмониях с интерстициальным типом инфильтрации, вызванных, например, вирусами, легионеллами, микоплазмами и т.д. Для таких интерстициальных пневмоний характерно преобладание снижения воздушности по типу «матового стекла» над консолидацией [13, 14].

В Российской Федерации приняты Временные рекомендации, согласно которым в заключении рентгенолога должны быть указаны риск коронавирусной пневмонии, ее распространенность по четырем степеням и процент поражения легочной ткани [6]. В таких условиях от фак-

ORIGINAL STUDY

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

тически эмпирического, без применения объективных измерений заключения, сделанного рентгенологом, могут зависеть установление диагноза, стационарное или амбулаторное лечение, тактика ведения пациента и т.д. Несмотря на то что существуют инструменты по полуавтоматическому или автоматическому детектированию объема поражения легочной ткани, они неунифицированы, используют принципы порогового и нейросетевого анализа. Кроме того, законодательно их использование не может быть изолированно, без верификации решением сертифицированного врача-рентгенолога.

Для классификаций риска коронавирусной пневмонии было проведено несколько исследований по оценке сходимости результатов. Так, в исследовании M.O. Hadied et al. [7] была представлена оценка 70 KT-исследований органов грудной клетки тремя группами экспертов для выделения одного из четырех вариантов заключения: характерные и нехарактерные, неопределенные признаки и признаки иного заболевания. В каждой группе наблюдалось умеренное согласие между наблюдателями, коэффициент согласованности Каппа составил  $0.45-0.52 \pm 0.02$ ;  $0.58 \pm 0.06$  и  $0.61 - 0.67 \pm 0.06$  соответсвенно. В исследовании Т.М.Н. DeJaegere et al. [8] для аналогичной классификации риска оценка производилась между двумя врачами-рентгенологами с опытом работы менее 5 лет и опытным врачом-рентгенологом. Коэффициент согласованности среди менее опытных наблюдателей составил 0,633, при этом согласие между наблюдателями разного уровня составило 0,570, что соответствует умеренному согласию между наблюдателями. В работе N. Sushentsev et al. [9] производилась внутри- и межэкспертная оценка согласия при применении систем отчетности COVID-RADS и CO-RADS среди рентгенологов с разным опытом, которая оценивалась с использованием коэффициентов Флейсса и Коэна. Значения коэффициента по Флейссу составили 0,74 и 0,73 для COVID-RADS и CO-RADS соответственно. По коэффициенту Коэна отмечалось умеренное согласие между наблюдателями, равное 0,51.

Система с оценкой распространенности по КТ-степеням с указанием процента поражения внедрена только в Российской Федерации, в связи с этим публикаций по сходимости результатов подобной экспертной оценки нами найдено не было.

**Цель исследования** — определить меж- и внутриэкспертную сходимость в оценке процента поражения и КТ-степени COVID-19-ассоциированной пневмонии при КТ органов грудной клетки среди экспертов с разным опытом, проанализировать уровень сходимости в зависимости от распространенности поражения и иных факторов.

# Материалы и методы

В исследовании анализировались данные 221 пациента, которые были госпитализированы в самарский COVID-госпиталь Клиник СамГМУ с апреля по декабрь 2020 г. с подтвержденным ПЦР SARS-CoV-2. Всем пациентам

#### S.S. Pervushkin, P.M. Zelter, E.K. Kramm, E.A. Sartakova

Samara State Medical University, Samara, Russian Federation

# Comparison of Intra- and Inter-Observer Consistency (Intra-Expert Reliability, Inter-Expert Reliability) in Assessing the Extent of COVID-19 Pneumonia Lesions on Chest Computed Tomography

Background. The SARS-CoV-2 pandemic has set new challenges for the radiological community: early diagnosis of interstitial pneumonia, estimation of its severity, and dynamics. Computed tomography has become the method of choice for assessing lung tissue in COVID-19 patients, which is due to the low sensitivity of radiography in detecting a decrease in airiness by the "ground glass opacity" type. The criteria for assessing visual signs of lung tissue damage often have a degree of subjectivity, and the conclusion based on them affects the patient's management tactics. Aims — to determine the inter- and intra-expert consistency in the assessment of the percentage of lesions and the CT stage of COVID-19 pneumonia among experts-radiologists with different experience, to analyze the level of precision depending on the prevalence of the lesion and other factors. Materials and methods. The research analyzed CT scans of 221 patients with confirmed SARS-CoV-2 by PCR. Patients with additional lung pathology and some patients with lesions up to 50% were excluded to create a uniform degree sample of 100 patients. Four expert radiologists determined the percentage of lung damage and the CT stages. The results of the expert assessment are analyzed using the methods of classical descriptive statistics and the analysis of intra-and inter-expert consistency. Results. The correlation of the level of lung damage, when evaluating intra-expert convergence (after 6 months), as a percentage between the first and second reading was R = 0.86 (p < 0.05) for expert 1 (high level of training), R = 0.84(p < 0.05) for expert 2 (high level of training). Within the expert agreement, Kappa (for K-degree) was 0.54 for expert 1 and 0.46 for expert 2, which corresponds to a moderate level of consistency. When assessing inter-expert convergence, the connectivity between the level of lung damage as a percentage between experts 1 and 2 (high level of training) was R = 0.87 (p < 0.05), between experts 3 and 4 (low level of training) -R = 0.78(p < 0.05). The measure of inter-expert agreement Kappa was 0.51 for experts 1 and 2 and 0.56 for experts 3 and 4. The average assessment of experts in the sample varied up to 4.5%, and when analyzing the differences in expert opinions, the difference varied evenly, both in the degree of increase and decrease in the volume of the lesion. Conclusions. In spite of the high level of correlation in the assessment of the percentage of lesion, the convergence of the Cap on the CT degree was moderate, not significantly differing from the degree of expert training. More often, differences in the level of damage are observed in "non-classical" patterns: "reverse halo", curvilinear lines, etc. The difference in opinions doesn't represent a systematic error. Hereby the expert assessment of the volume of lung damage "empirically" has a moderate, insufficient level of reliability.

 $\textbf{\textit{Keywords}: Coronavirus Infections, COVID-19, SARS-CoV-2, virial pneumonia, intra-expert reliability, inter-expert reliability}$ 

For citation: Pervushkin SS, Zelter PM, Kramm EK, Sartakova EA. Comparison of Intra- and Inter-Observer Consistency (Intra-Expert Reliability, Inter-Expert Reliability) in Assessing the Extent of COVID-19 Pneumonia Lesions on Chest Computed Tomography. Annals of the Russian Academy of Medical Sciences. 2022;77(2):97–106. doi: https://doi.org/10.15690/vramn1644

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

была проведена КТ грудной клетки на сканерах Toshiba Aquilion 32 (Япония) и GE Revolution EVO 128 (США). Томография проводилась в рутинном режиме при напряжении на рентгеновской трубке 120 кВ и автоматическом расчете параметров силы тока. Произведена первичная оценка объема поражения и КТ-степени, оформленная в виде стандартного протокола в соответствии с Временными методическими рекомендациями [6]. В нативном распределении превалировали случаи КТ-1 (64%) и КТ-2 (17%), далее с целью более равномерного распределения и сокращения выборки до ровного числа исключен 121 пациент со стадиями КТ-1 и КТ-2. Таким образом, сформирована выборка серий КТ грудной клетки в количестве 100 штук с распределением по КТ-степеням: KT-1 (33%), KT-2 (25%), KT-3 (28%) и KT-4 (14%), Исследования были анонимизированы и рандомизированы с присвоением номера. Полученная выборка КТ была представлена четырьмя экспертам с различным уровнем подготовки, которым была поставлена задача определить процент поражения легких и КТ-степень в соответствии с Временными методическими рекомендациями [6].

Описание уровня подготовки экспертов:

 эксперт 1 — врач-рентгенолог стационара с опытом работы в торакальной радиологии 5 лет;

- эксперт 2 врач-рентгенолог стационара с опытом работы в торакальной радиологии 8 лет;
- эксперт 3 врач-ординатор по специальности «рентгенология» с опытом работы 6 мес;
- эксперт 4 —врач-ординатор по специальности «рентгенология» без опыта работы, получивший инструкцию

Первое чтение для оценки внутриэкспертной сходимости производилось с использованием ретроспективных данных протоколов описания КТ при поступлении пациентов (эксперт 1, n = 55; эксперт 2, n = 31). Второе чтение производилось из выборки КТ-исследований, включающей КТ первого чтения. Эксперты 3 и 4 просматривали подборку однократно. Общая схема исследования представлена на рис. 1.

Произведен анализ результатов экспертной оценки методами классической описательной статистики, определялся коэффициент корреляции Пирсона между результатами оценки поражения в процентах и анализировались диаграммы рассеяния.

Экспертная сходимость оценивалась в программе SPSS Statistics (IBM, США). Коэффициент каппа Коэна использовался для измерения внутри- и межэкспертной сходимости для качественных (категориальных) пунктов:

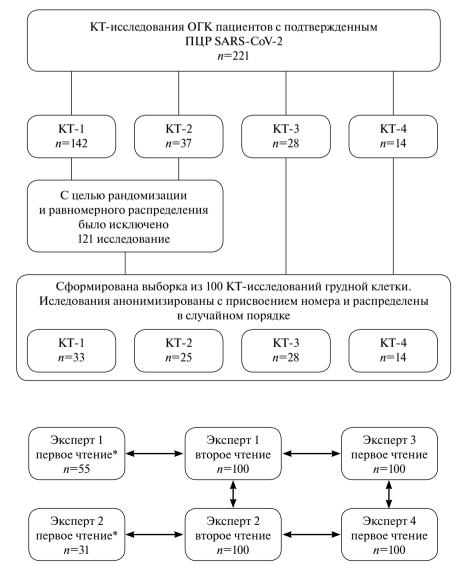


Рис. 1. Схема исслелования

KT-степеней, а также анализ комбинаторных таблиц сравнения экспертных оценок.

Экспертная сходимость в контексте статьи разделена на внутри- и межэкспертную. В случае внутриэкспертной сходимости сравниваются оценки одного и того же эксперта, произведенные в разное время, представляющие собой аналог воспроизводимости. В свою очередь, межэкспертная сходимость представляет собой оценку несколькими экспертами одного и того же объекта. Важность экспертной сходимости заключается в том, что она отражает степень, в которой данные, собранные в ходе исследования, являются правильным представлением измеряемых переменных, особенно при отсутствии возможности сравнения с эталоном, или «золотым стандартом». Измерение степени, в которой эксперты присваивают одинаковые баллы одному и тому же событию, называется экспертной сходимостью.

Традиционно используется процент согласия, рассчитываемый как отношение количества оценок согласия к общему количеству оценок. Однако J. Cohen, критикуя использование процентного соглашения, представил каппу Коэна (и) как более надежную статистическую меру, чем простой расчет процента согласия, поскольку при этом подходе учитывается возможность случайных совпадений [16].

Внутриэкспертная сходимость представляет собой сравнение результатов оценки одного и того же эксперта в различных условиях или по прошествии некоторого времени, в некоторой степени напоминая воспроизволимость.

Произведены статистический анализ разницы результатов экспертной оценки и оценка уровня частот расхождения мнений экспертов в уровне процента расхождения. Для того чтобы понимать причины расхождения в мнениях экспертов, размер и направление (в сторону занижения объема или его увеличения) этих расхождений, мы рассмотрели разницу в оценке как отдельный статистический элемент. Для получения разницы (Рн) мы отнимали значение оценки одного эксперта от значения другого:  $PH = 9 \kappa c \pi 1(\%) -$ -Эксп2(%). Таким образом, Рн могла принимать как отрицательные, так и положительные значения и в случае близких результатов стремиться к нулю. Для наглядного представления построены частотные диаграммы, где на оси абсцисс представлена разница в мнениях, а на оси ординат — количество случаев. Соответственно, при систематической ошибке фигура диаграммы будет смещена в сторону от нуля, а при случайной — представлять фигуру, близкую к равнобедренному треугольнику с вершиной в точке 0, при этом чем уже основание и острее вершина, тем меньше размах и частота ошибки, и наоборот.

#### Результаты

В первую очередь все полученные оценки экспертов в виде процента поражения были проверены по одновыборочному критерию Колмогорова-Смирнова на предмет нормальности распределения, а вычисление средней оценки представлено в табл.1. Минимальный критерий Колмогорова—Смирнова составил 0.104, следовательно. в проверяемых наборах данных определяется нормальное распределение. При анализе средних оценок наибольшая разница показана при первом чтении, средняя оценка эксперта 2 была на 4,5% выше оценки эксперта 1, при этом в данном случае оценивалась неполная выборка. При оценке полной выборки из 100 пациентов наибольшая разница в средней оценке наблюдалась между экспертами 3 и 4 (низкий уровень подготовки), составив 2,14%. Таким образом, можно заключить, что расхождения экспертов в целом имеют характер случайной ошибки

Первая оценка экспертов 1 и 2 получена ретроспективно при анализе протоколов описания КТ органов грудной полости, выполненных при поступлении пациентов.

На рис. 2 представлены диаграммы рассеяния экспертной оценки объема поражения при первом и втором чтении. Корреляция уровня поражения легких в процентах между первым и вторым чтением составила у эксперта 1 R=0.86 (p<0.05), у эксперта 2-R=0.84 (p<0.05). Таким образом, эксперты продемонстрировали высокую и сопоставимую внутриэкспертную корреляцию оценок уровня поражения.

Максимальная разница между первым и вторым просмотром в оценке процента поражения у эксперта 1 составила 30%, у эксперта 2-27,5%, при этом среднее значение разницы составило 7,8 и 9,6% соответственно. На рис. 3 представлены частотные гистограммы внутриэкспертной разницы в процентах.

Оценка внутриэкспертной сходимости производилась исходя из КТ-степени, установленной по уровню поражения легких. Из 55 пациентов, просмотренных дважды экспертом 1, количество совпадений по степени составило 37 (67,2%). Эксперт 2 оценил 31 пациента, количество совпадений — 19 (61,2%). Разногласия находились в пределах 1—3-й степени (табл. 2).

Таблица 1. Одновыборочный критерий Колмогорова-Смирнова

			№ эксперта						
	2	1	1	2	2				
		Первое чтение Вто		Второе	чтение	3	4		
N		55	31	100	100	10	100		
Распределение *,**	Среднее	36,56	41,03	38,22	38,32	37,54	39,68		
	Среднеквадратичные отклонения	23,49	24,55	24,26	24,45	27,19	28,19		
	Абсолютные	0,180	0,157	0,104	0,127	0,160	0,169		
Отклонения	Положительные	0,180	0,157	0,104	0,127	0,160	0,169		
	Отрицательные	-0,093	-0,103	-0,071	-0,082	-0,099	-0,106		
Статистика критерия		0,180	0,157	0,104	0,127	0,160	0,169		

<sup>\*</sup> Проверяемое распределение является нормальным.

<sup>\*\*</sup> Вычислено из данных.

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

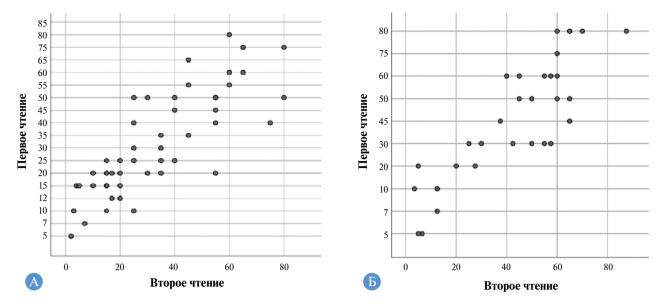


Рис. 2. Диаграмма значений процента поражения экспертов 1 (А) и 2 (Б)

Мера внутриэкспертного согласия Каппа составила 0,54 для эксперта 1 и 0,46 для эксперта 2, что соответствует умеренному (moderate) уровню согласованности [10].

Для оценки межэкспертной сходимости использовались результаты второго чтения, количество пациентов 100. Описательная статистика результатов второго чтения экспертами представлена в табл. 3: среднее значение процента поражения легких у опытных экспертов (1 и 2) практически идентичное, что, скорее всего, объясняется

«более спокойными условиями» оценки в сравнении с первым чтением.

Разница в среднем уровне межэкспертной оценки у менее опытных экспертов (3 и 4) составила 2%, второе чтение не проводилось.

Взаимосвязь между оценкой уровней поражения легких в процентах между экспертами 1 и 2 составила R=0.87 (p<0.05), между экспертами 3 и 4 — R=0.78 (p<0.05). На рис. 4 представлены диаграммы рассеяния межэкспертной оценки объема поражения экспер-

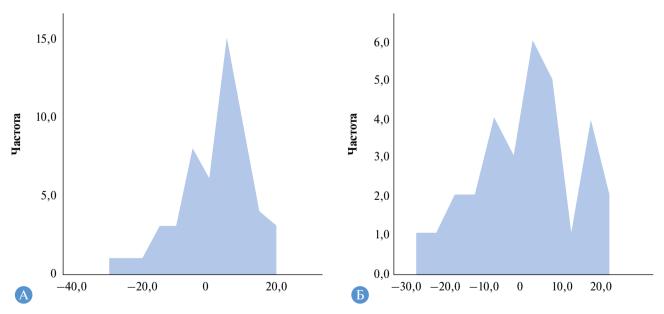


Рис. 3. Частотные гистограммы внутриэкспертной разницы экспертов 1 (А) и 2 (Б), %

Таблица 2. Комбинационная таблица внутриэкспертного согласия, эксперты 1 и 2

		Э:	Эксперт 1, второе чтение			
		1	2	3	4	
Эксперт 1, первое чтение	1	18	3	1	0	
	2	2	9	2	1	
	3	0	5	6	1	
	4	0	0	3	4	

		Эксперт 2, второе чтение				
		1	2	3	4	
Эксперт 2, первое чтение	1	8	1	0	0	
	2	0	4	4	0	
	3	0	3	6	0	
	4	0	0	4	1	

Таблица 3. Описательная статистика результатов второго чтения

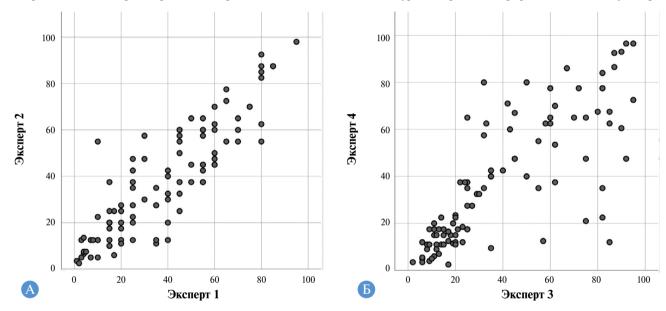
Эксперт	N	Минимум	Максимум	Среднее
1	100	1,0	95,0	38,32
2	100	2,5	98,0	38,22
3	100	2,5	96,5	37,54
4	100	2,0	95,0	39,68

тами высокого и низкого уровней подготовки. Обращает на себя внимание, что расхождение в оценке у экспертов низкого уровня более выражено при объеме поражения 40% и более.

При анализе разницы в межэкспертной оценке наблюдалось, что у экспертов высокого уровня подготовки максимальное расхождение в оценке составило 27,5% (среднее -9,3%), а у экспертов низкого уровня -59,5% (среднее -11,6%). На рис. 5 представлены частотные гистограммы межэкспертной разницы в процентах.

Оценка межэкспертной сходимости исходя из КТ-степени продемонстрировала: из 100 пациентов у экспертов 1 и 2 количество совпадений по степени составило 65 (65%), у экспертов 3 и 4 — 69 (69%). Разногласия находились в пределах 1—3-й степени (табл. 4). Мера межэкспертного согласия Каппа составила 0,51 для экспертов 1 и 2 и 0,56 для экспертов 3 и 4, что соответствует умеренному (moderate) уровню согласованности [10].

Тем не менее, несмотря на то что оценщики менее опытного уровня продемонстрировали более слабую кор-



**Рис. 4.** Диаграмма значений процента поражения между экспертами: A-1 и 2 (высокий уровень подготовки); B-3 и 4 (низкий уровень подготовки)

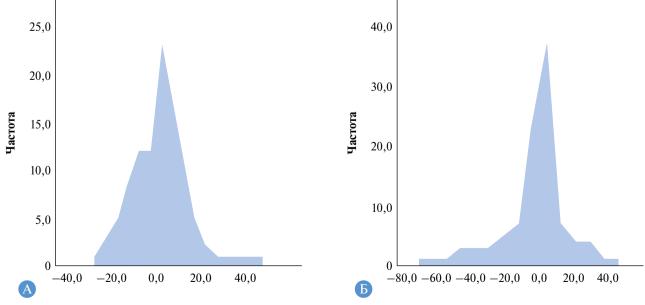


Рис. 5. Частотные гистограммы межэкспертной разницы в процентах: А — высокий уровень подготовки; Б — низкий уровень подготовки

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

Таблица 4. Комбинационная таблица межэкспертного согласия, эксперты высокого (1 и 2) и низкого (3 и 4) уровней

		Э	Эксперт 2, второе чтение			
		1	2	3	4	
Эксперт 1, второе чтение	1	28	5	1	0	
	2	7	17	6	0	
	3	0	10	13	2	
	4	0	0	4	7	

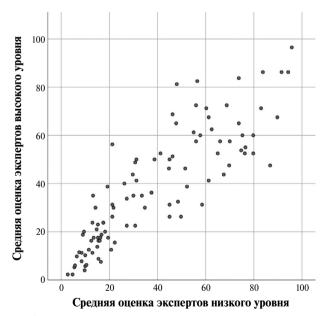
реляцию между своими оценками объема поражения в процентах, количество совпадений по KT-степени оказалось немного большим.

Для оценки взаимосвязи экспертного мнения врачей низкого и высокого уровней подготовки было рассчитано среднее арифметическое уровня поражения легких в процентах у каждой из групп экспертов и переведены средние значения в КТ-стадию. Корреляция среднего процента составила R=0.86 (p<0.05). На рис. 6 представлена диаграмма рассеяния средней оценки объема поражения экспертами высокого и низкого уровней подготовки.

Оценка межэкспертной сходимости исходя из КТ-степени показала, что из 100 пациентов количество совпадений по степени составило 67 (67%). Мера межэкспертного согласия Каппа — 0,53. Разногласия находились в пределах степени 1-3 (табл. 5).

В качестве примеров расхождения при межэкпертной оценке представляем два клинических случая.

Пациентка Т., 57 лет, поступила в ковидный госпиталь Клиник СамГМУ с жалобами на умеренную общую слабость, частый сухой кашель в течение дня, повы-



**Рис. 6.** Диаграмма значений среднего процента поражения между экспертами разного уровня подготовки

**Таблица 5.** Комбинационная таблица согласия между экспертами разного уровня подготовки

		Среднее	, высокий у	ровень подготовки		
		1	2	3	4	
Среднее, низкий уровень подготовки	1	35	6	1	0	
	2	2	15	7	2	
	3	0	6	13	2	
	4	0	1	7	4	

			Эксперт 4			
		1	2	3	4	
Эксперт 3	1	39	3	1	3	
	2	2	14	3	3	
	3	0	6	7	5	
	4	0	1	4	9	

шение температуры до 38,5 °C (макс), головную боль, сухость во рту, скудные слизистые выделения из носа, заложенность носа, небольшую одышку при физической нагрузке. При поступлении была выполнена КТ органов грудной клетки, сканы в разных плоскостях представлены на рис. 7.

В обоих легких с преимущественно субплевральным распространением определяются зоны снижения воздушности легочной ткани по типу «матового стекла» с неровными, местами нечеткими контурами, имеющими тенденцию к слиянию.

Пациент Б., 48 лет, поступил в ковидный госпиталь Клиник СамГМУ с жалобами на общую слабость, головную боль в лобной области, нарушение вкуса и обоняния, повышенную температуру тела до 38 °C, озноб, одышку, затруднение дыхания, кашель с трудноотделяемой прозрачной мокротой на протяжении всего дня. В сортировочном пункте выполнена КТ органов грудной клетки, и с заключением «Двусторонняя интерстициальная пневмония КТ-3 (до 60%)» пациент был госпитализирован. На 11-й день госпитализации в условиях Клиник СамГМУ выполнена КТ органов грудной клетки (сканы представлены на рис. 8).

На представленных сканах определяются рассеянные зоны криволинейного уплотнения легочной ткани, локализованного преимущественно в субплевральных зонах и в периферических отделах легочных сегментов.

Экспертами была оценена площадь поражения легочной ткани (табл. 6).

При анализе представленных случаев обращает на себя внимание выраженное расхождение в оценке пораженного легочного интерстиция, что может говорить о низком уровне межэкспертной сходимости при оценке объема поражения, напрямую зависящем от опыта эксперта, знания клинической картины конкретного пациента, его анамнеза. Различия также наблюдаются чаще при «неклассических» паттернах: «обратном гало», криволинейных уплотнениях и т.д. Если в случае с пациенткой Т. расхождение некритическое и не сильно влияет на тактику лечения, то в случае с пациентом Б. это может иметь выраженное клиническое значение: при оценке поражения в 85% пациент может быть неправильно отнесен к негативной прогностической группе, а при оценке 10%, наоборот, может произойти недооценка его состояния, в данных случаях напрямую влияя на тактику ведения пациентов, предварительный прогноз, а также на длительность пребывания пациентов в стационаре. Помимо знания клинической картины и представления о динамике развития патологического процесса, на более точную оценку объема поражения легочной ткани влияет также «насмотренность» экспертов 1 и 2 в сравнении с менее опытными экспертами 3 и 4. Немаловажное значение, на наш взгляд, имеет и фактор усталости и «замыленности» экспертов, в какое время была проведена оценка предложенных клинических случаев, так как часть экс-

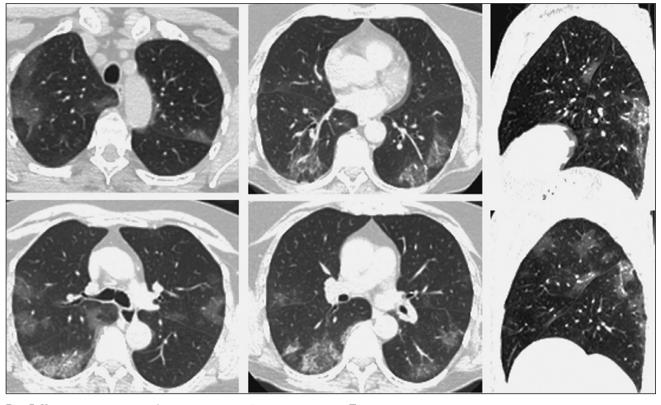


Рис. 7. Компьютерная томография органов грудной клетки пациентки Т.

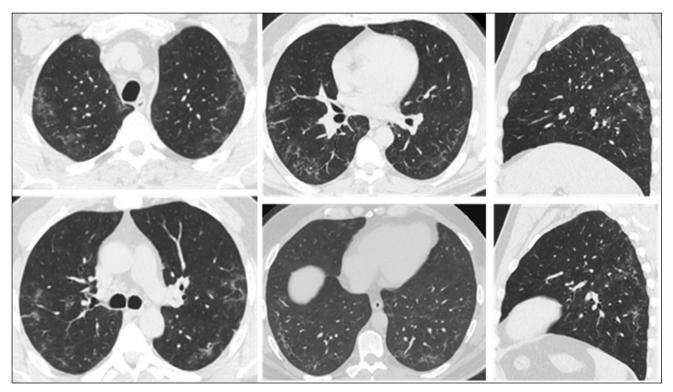


Рис. 8. Компьютерная томография органов грудной клетки пациента Б.

**Таблица 6.** Результаты оценки уровня поражения в представленных клинических примерах, %

	Эксперт						
	1	2	3	4			
Пациентка Т.	25 (KT-2)	12,5 (KT-1)	17,5 (KT-1)	25 (KT-2)			
Пациент Б.	10 (KT-1)	55 (KT-3)	12 (KT-1)	85 (KT-4)			

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

пертов работала в выходной день, а часть — в конце рабочего дня (об этом может свидетельствовать расхождение в оценке у более опытных экспертов 1 и 2).

## Обсуждение

Оценка сходимости при экспертной оценке применяется достаточно широко в медицинской науке вообще и лучевой диагностике в частности. Тактика ведения пациентов с установленной или предполагаемой COVID-19-ассоциированной пневмонией подробно изложена во Временных методических рекомендациях 13-й версии. Так, решение о необходимости госпитализации пациента принимается врачом на основании комплекса клиникоэпидемиологических данных с учетом тяжести состояния пациента (среднетяжелое/тяжелое течение заболевания) и требований, предусмотренных приказом Минздрава России от 19.03.2020 № 198н. К таким требованиям относят выполнение следующих критериев: SpO<sub>2</sub> < 95%,  $T \ge 38$  °C и ЧДД > 22. Результаты КТ не фигурируют в данном приказе и не должны быть основанием для выбора стационарного или амбулаторного лечения [17]. В Методических рекомендациях 13-й версии по сравнению с предыдущими рекомендовано не использовать термин «КТ-тяжесть» в связи с тем, что тяжесть состояния пациента не всегда соответствует распространению поражения по КТ. Несмотря на это, в разных регионах на фоне выраженного дефицита коечного фонда определение объема поражения легочной ткани по КТ считалось одним из критериев тяжести процесса и опосредованно являлось основанием для стратификации пациентов для госпитализации [18]. В связи с этим нами и был рассмотрен вопрос о сходимости определения объема поражения при оценке одним или несколькими врачами. На наш взгляд, определение степени поражения более целесообразно для прогнозирования исхода заболевания и оценки динамики процесса, как показано в работе С.П. Морозова и соавторов [19].

Оценка диагностических изображений предполагает определенную степень субъективизма, в особенности при необходимости оценки распространенности патологического процесса. Применительно к пневмонии, вызванной SARS-CoV-2, межэкспертная сходимость анализировалась относительно шкал вероятности CoRADS [4] и COVID-RADS [9]. Шкала процентной оценки поражения легких не имеет широкого распространения в мире, но применяется в России согласно Временным рекомендациям [6]. В связи с этим важный вопрос — проверка повторяемости врачебного подсчета поражения легких. Мы изучили сходимость при первичной и повторной оценке рандомизированной группы пациентов. Выявлено, что у экспертов с большим опытом в торакальной радиологии наблюдался более высокий уровень (R - 0.86и 0,84) по сравнению с экспертами с небольшим опытом (R-0.54 и 0.46). При оценке межэкспертной сходимости Каппа составила 0,51 для более опытных экспертов и 0,56 для менее опытных. Данные результаты иллюстрируют высокий уровень субъективности при оценке, значимость местных принятых правил расчета, которые могут варьировать в зависимости от лечебного учреждения. В таких условиях встает вопрос, насколько целесообразно оценивать процент поражения и стадию, насколько результаты анализа воспроизводимы. Описанная проблема может быть решена внедрением в клиническую практику инструментов для автоматической оценки легочной паренхимы [11]. В таких инструментах, основанных на работе искусственного интеллекта, могут наблюдаться ошибки выделения, которые незначительно влияют на общий процент, но по сравнению с визуальным анализом отсутствует субъективность [12].

#### Заключение

Таким образом, в результате проведенного анализа было показано, что оценка поражения легочной ткани при вирусной пневмонии различными врачами обладает умеренной сходимостью, что может являться недостаточным для применения эмпирической визуальной оценки для достоверной стратификации по степеням поражения или оценки динамики процесса при контрольных исследованиях. Выявлено, что уровень ошибок имеет прямую зависимость от объема поражения, т.е. при большой распространенности патологического процесса врачу труднее найти объективные критерии для постановки той или иной степени или процента поражения легочной ткани. В связи с этим актуальной темой для изучения станет использование систем на основе искусственного интеллекта для сегментации и определения объема поражения легочной ткани, которое не зависит от опыта врача-рентгенолога.

### Дополнительная информация

**Источник финансирования.** Исследования выполнены, рукопись подготовлена и публикуется за счет финансирования по месту работы авторов.

**Конфликт интересов.** Авторы сообщают об отсутствии явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Участие авторов. С.С. Первушкин — анализ и интерпретация результатов исследования, написание статьи, подготовка статьи к публикации; П.М. Зельтер — концепция и дизайн исследования, руководство исследованием, проверка статьи перед публикацией; Е.К. Крамм — редактирование текста статьи, участие в написании статьи; Е.А. Сартакова — сбор клинических данных, анализ медицинской литературы. Все авторы внесли существенный вклад в проведение исследования и подготовку статьи, прочли и одобрили финальную версию до публикации.

# ЛИТЕРАТУРА

- Kim H. Outbreak of novel coronavirus (COVID-19): What is the role of radiologists? *Eur Radiol*. 2020;30(6):3266-3267. doi: https://doi.org/10.1007/s00330-020-06748-2
- Fang Y, Zhang H, Xie J, et al. Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR. *Radiology*. 2020;296(2):E115–E117. doi: https://doi.org/10.1148/radiol.2020200432
- Salehi S, Abedi A, Balakrishnan S, et al. Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. AJR Am J Roentgenol. 2020;215(1):87–93. doi: https://doi.org/10.2214/AJR.20.23034
- Prokop M, van Everdingen W, van Rees Vellinga T, et al. CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having

**ORIGINAL STUDY** 

Annals of the Russian Academy of Medical Sciences. 2022;77(2):97-106.

- COVID-19-Definition and Evaluation. *Radiology*. 2020;296(2):E97–E104. doi: https://doi.org/10.1148/radiol.2020201473
- Salehi S, Abedi A, Balakrishnan S, et al. Coronavirus disease 2019 (COVID-19) imaging reporting and data system (COVID-RADS) and common lexicon: a proposal based on the imaging data of 37 studies. *Eur Radiol*. 2020;30(9):4930–4942. doi: https://doi.org/10.1007/s00330-020-06863-0
- 6. Авдеев С.Н., Адамян Л.В., Алексеева Е.И., и др. Профилактика, диагностика и лечение новой коронавирусной инфекции (COVID-19): Временные методические рекомендации/Министерство здравоохранения РФ; утверждено зам. министра здравоохранения РФ Е.Г. Камкин. — М., 2020. — 236 с. [AvdeevNS, AdamjanLV, AlekseevaEI, i dr. Profilaktika, diagnostika I lechenie novoj koronavirusnoj infekcii (COVID-19): Vremennye metodicheskie rekomendacii / Ministerstvo zdravoohranenija RF; utverzhdeno zam. Ministra zdravoohranenija RF E.G. Kamkin. Moscow; 2020. 236 s. (In Russ.)]
- Hadied MO, Patel PY, Cormier P, et al. Interobserver and Intraobserver Variability in the CT Assessment of COVID-19 Based on RSNA Consensus Classification Categories. *Acad Radiol.* 2020;27(11):1499–1506. doi: https://doi.org/10.1016/j.acra.2020.08.038
- De Jaegere TMH, Krdzalic J, Fasen BACM, et al. Radiological Society of North America Chest CT Classification System for Reporting COVID-19 Pneumonia: Interobserver Variability and Correlation with Reverse-Transcription Polymerase Chain Reaction. *Radiol Cardiothorac Imaging*. 2020;2(3):e200213. doi: https://doi.org/10.1148/ryct.2020200213
- Sushentsev N, Bura V, Kotnik M, et al. A head-to-head comparison of the intra- and interobserver agreement of COVID-RADS and CO-RADS grading systems in a population with high estimated prevalence of COVID-19. *BJR Open*. 2020;2(1):20200053. doi: https://doi.org/10.1259/bjro.20200053
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977:159–174.
- Harmon SA, Sanford TH, Xu S, et al. Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets. *Nat Commun.* 2020;11(1):4080. doi: https://doi.org/10.1038/s41467-020-17971-2
- Adly AS, Adly AS, Adly MS. Approaches Based on Artificial Intelligence and the Internet of Intelligent Things to Prevent the Spread of

- COVID-19: Scoping Review. *J Med Internet Res.* 2020;22(8):e19104. doi: https://doi.org/10.2196/19104
- Zhao W, Zhong Z, Xie X, et al. Relation between Chest CT Findings and Clinical Conditions of Coronavirus Disease (COVID-19) Pneumonia: A Multicenter Study. AJR Am J Roentgenol. 2020;214(5):1072-1077. doi: https://doi.org/10.2214/ajr.20.22976
- Kumar V, Abbas AK, Fausto N, etal. Robbins and Cotran Pathologic Basis of Disease (Robbins Pathology). Saunders. ISBN:B005WV2Q86.
- 15. Зайратьянц О.В., Самсонова М.В., Михалева Л.М., и др. Патологическая анатомия COVID-19: Атлас / под общ. ред. О.В. Зайратьянца. М.: НИИОЗММ, 2020. 140 с. [Zajrat'yanc OV, Camsonova MV, Mihaleva LM, i dr. Patologicheskaya anatomiya COVID-19: Atlas / pod obshch. red. O.V. Zajrat'yanca. Moscow: NIIOZMM; 2020. 140 s. (In Russ.)]
- McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012;22(3):276–282.
- Порядок госпитализации в медицинские организации пациентов с установленным диагнозом новой коронавирусной инфекции COVID-19 в зависимости от степени тяжести заболевания: Приложение № 12 к приказу Минздрава России от 19.03.2020 № 198н. Available from: http://www.consultant.ru/document/cons\_doc\_LAW\_348101/113f06199319dbb50e28daaaac 80e34214450766/
- Эволюция пандемии COVID-19 / под ред. Н.А. Белякова, С.Ф. Багненко. — СПб.: Балтийский медицинский образовательный центр, 2021. — 411 с. [Evolyuciya pandemii COVID-19 / pod red. N.A. Belyakova, S.F. Bagnenko. Saint Petersburg: Baltijskij medicinskij obrazovatel'nyj centr; 2021. 411 s. (In Russ.)]
- 19. Морозов С.П., Чернина В.Ю., Блохин И.А., и др. Прогнозирование исходов при лабораторно верифицированном COVID-19 по данным компьютерной томографии органов грудной клетки: ретроспективный анализ 38 051 пациента // Digital Diagnostics. 2020. Т. 1. № 1. С. 27—36. [Morozov SP, Chernina VYu, Blokhin IA, Gombolevskiy VA. Chest computed tomography for outcome prediction in laboratory-confirmed COVID-19: Aretrospective analysis of 38,051 cases. *Digital Diagnostics*. 2020;1(1):27—36. (In Russ.)] doi: https://doi.org/10.17816/DD46791

# КОНТАКТНАЯ ИНФОРМАЦИЯ

**Зельтер Павел Михайлович**, к.м.н., доцент [*Pavel M. Zelter*, MD, PhD, Associate Professor]; **адрес**: 443099, Самара, ул. Чапаевская, д. 89 [**address**: 89, Chapaevskaya str., 443099, Samara, Russia]; **e-mail**: pzelter@mail.ru, **SPIN-ко**д: 3678-3932, **ORCID**: https://orcid.org/0000-0003-1346-5942

Первушкин Сергей Сергеевич, ассистент [Sergey S. Pervushkin, Assistant]; e-mail: sergey.pervushkin@gmail.com, SPIN-код: 3089-5082, ORCID: https://orcid.org/0000-0002-7574-283X

*Крамм Евгения Константиновна*, ассистент [*Evgeniya K. Kramm*, Assistant]; **e-mail**: Evgeniyakramm@gmail.com, SPIN-код: 4826-5241, ORCID: https://orcid.org/0000-0003-3029-8787

**Сартакова Елизавета Андреевна**, клинический ординатор [*Elizaveta A. Sartakova*, Resident]; e-mail: Sartakova163@mail.ru, SPIN-код: 6825-5077, ORCID: https://orcid.org/0000-0002-2439-197X