

# Воспроизводимость цитологических диагнозов при оценке жидкостных мазков шейки матки, а также иммуноцитохимической коэкспрессии p16/Ki-67 ручным и автоматическим методом

А.В. Трегубова, Н.С. Теврюкова, Л.С. Ежова, М.В. Шамаракова, А.С. Бадлаева, Д.А. Добровольская, Г.Р. Байрамова, Н.М. Назарова, А.Ю. Шилиев, А.В. Асатурова✉

ФГБУ «Национальный медицинский исследовательский центр акушерства, гинекологии и перинатологии им. акад. В.И. Кулакова» Минздрава России, Москва, Россия

## Аннотация

Цель. Оценить воспроизводимость цитологических диагнозов при оценке жидкостных мазков шейки матки, а также иммуноцитохимической коэкспрессии p16/Ki67 ручным и автоматическим методом.

Материалы и методы. Исследовались цитологические мазки, приготовленные по методу жидкостной цитологии на приборе Becton Dickinson (технология SurePath), иммуноцитохимическое исследование проводилось на автоматическом иммуноштейнере Ventana BenchMark Ultra с использованием коммерческого набора CINtec (определение коэкспрессии p16/Ki-67). Изучено 100 цитологических препаратов (50 пар ПАП-мазок-иммуноцитохимический препарат). Диагностический набор просмотрен пятью цитологами независимо друг от друга, цитологические препараты оценены с помощью 4 категорий по системе Бетесда (2014 г.) и по категориям норма/патология. Оценка коэкспрессии p16/Ki-67 проводилась по рекомендациям производителя (Roch) ручным методом (световой микроскоп) и с помощью автоматической системы Vision Cyto Pap ICC. Статистическая обработка полученных результатов проводилась с помощью пакета программного обеспечения SPSS версии 26.0.0.0 с определением показателей воспроизводимости каппа Коэна и каппа Флейса.

Результаты. При оценке воспроизводимости четырех категорий цитологических диагнозов по системе Бетесда (2014 г.) выявлено, что уровень коэффициента каппа Коэна составил 0,048–0,265. Общая каппа Флейса между всеми цитологами составила 0,103. При выделении только двух категорий (норма и патология) уровень воспроизводимости составил от 0,058 до 0,377. При оценке коэкспрессии p16 и Ki-67 выявлена воспроизводимость по каппе Коэна от 0,196 до 0,574, при этом общая каппа Флейса составила 0,407. При сопоставлении результатов оценки каждым из цитологов с нейросетью воспроизводимость по каппе Коэна составила от 0,103 до 0,436.

Заключение. В настоящем исследовании воспроизводимость цитологических заключений с применением системы Бетесда 2014 г., а также двух категорий (норма/патология) на основании исследования ПАП-мазков оказалась низкой. С нашей точки зрения, такие результаты прежде всего связаны с большим количеством патологических мазков, включенных в исследование. Применение иммуноцитохимического метода позволило в 3 раза увеличить воспроизводимость диагнозов, что свидетельствует о необходимости использования определения коэкспрессии p16 и Ki-67 для повышения чувствительности и специфичности цитологического метода. Сопоставимые данные воспроизводимости при сравнении ручной и автоматической оценки «двойной метки» свидетельствуют о том, что в настоящее время нейросетевой алгоритм может помочь в области поддержки принятия решения, а не заменить цитолога на диагностическом этапе.

**Ключевые слова:** ПАП-мазок, коэкспрессия p16/Ki-67, нейросетевой анализ, воспроизводимость, система Бетесда

**Для цитирования:** Трегубова А.В., Теврюкова Н.С., Ежова Л.С., Шамаракова М.В., Бадлаева А.С., Добровольская Д.А., Байрамова Г.Р., Назарова Н.М., Шилиев А.Ю., Асатурова А.В. Воспроизводимость цитологических диагнозов при оценке жидкостных мазков шейки матки, а также иммуноцитохимической коэкспрессии p16/Ki-67 ручным и автоматическим методом. Гинекология. 2022;24(6):499–505. DOI: 10.26442/20795696.2022.6.202009

© ООО «КОНСИЛИУМ МЕДИКУМ», 2022 г.

## Информация об авторах / Information about the authors

✉ **Асатурова Александра Вячеславовна** – д-р мед. наук, зав. 1-м патолого-анатомическим отд-нием ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». E-mail: a.asaturova@gmail.com; ORCID: 0000-0001-8739-5209

**Трегубова Анна Васильевна** – мл. науч. сотр. 1-го патологоанатомического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0003-4601-1330

**Теврюкова Надежда Сергеевна** – канд. биол. наук, ст. науч. сотр. 1-го патологоанатомического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0003-3305-8543

**Ежова Лариса Сергеевна** – канд. мед. наук, ст. науч. сотр. 1-го патолого-анатомического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0002-9804-8349

**Шамаракова Марина Викторовна** – канд. мед. наук, ст. науч. сотр. 1-го патологоанатомического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0002-0972-4350

**Бадлаева Алина Станиславовна** – мл. науч. сотр. 1-го патологоанатомического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0001-5223-9767

**Добровольская Дарья Алексеевна** – аспирант научно-поликлинического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0002-1409-9959

✉ **Aleksandra V. Asaturova** – D. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. E-mail: a.asaturova@gmail.com; ORCID: 0000-0001-8739-5209

**Anna V. Tregubova** – Res. Assist., Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0003-4601-1330

**Nadezda S. Tevrukova** – Cand. Sci. (Biol.), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0003-3305-8543

**Larisa S. Ezhova** – Cand. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0002-9804-8349

**Marina V. Shamarakova** – Cand. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0002-0972-4350

**Alina S. Badlaeva** – Res. Assist., Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0001-5223-9767

**Darya A. Dobrovolskaya** – Graduate Student, Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0002-1409-9959

# Reproducibility of cytological diagnoses in evaluating liquid cervical smears and immunocytochemical co-expression of p16/Ki-67 using manual and automatic methods

Anna V. Tregubova, Nadezda S. Tevrukova, Larisa S. Ezhova, Marina V. Shamarakova, Alina S. Badlaeva, Darya A. Dobrovolskaya, Guldana R. Bayramova, Niso M. Nazarova, Alexey Yu. Shilyaev, Aleksandra V. Asaturova✉

Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia

## Abstract

**Aim.** To assess the reproducibility of cytological diagnoses in evaluating liquid cervical smears and immunocytochemical co-expression of p16/Ki-67 using manual and automatic methods.

**Materials and methods.** Cytological smears prepared using the liquid cytology method on the Becton Dickinson device (SurePath technology) were studied. An immunocytochemical study was carried out using a Ventana BenchMark Ultra automatic immunostainer with a commercial CINtec kit (determination of p16/Ki-67 co-expression). In total, 100 cytological slides (50 pairs of Pap-smears and immunocytochemical slides) were studied. The diagnostic kit was reviewed by five cytologists independently, and the cytologic slides were evaluated using four categories according to the Bethesda system (2014) and according to the categories of normal/abnormal. The co-expression of p16/Ki-67 was assessed per the manufacturer's recommendations (Roche) using the manual method (light microscope) and the automatic Vision Cyto Pap ICC system. Statistical processing of the results was performed using the SPSS software package version 26.0.0.0 with the calculation of the reproducibility indices of Cohen's kappa and Fleiss' kappa.

**Results.** When assessing the reproducibility of four categories of cytological diagnoses according to the Bethesda system (2014), Cohen's kappa was 0.048–0.265. The overall Fleiss' kappa between all cytologists was 0.103. When only two categories (normal/abnormal) were used, the reproducibility ranged from 0.058 to 0.377. When assessing the co-expression of p16 and Ki-67, Cohen's kappa reproducibility was from 0.196 to 0.574, while the overall Fleiss' kappa was 0.407. When comparing the evaluation results of each of the cytologists with the neural network, Cohen's kappa reproducibility ranged from 0.103 to 0.436.

**Conclusion.** The reproducibility of cytological diagnoses according to the Bethesda system (2014) and two categories (normal/abnormal) based on the Pap smear study was low. Such results are primarily due to a large number of abnormal smears in the study. The immunocytochemical method has diagnosis reproducibility three times higher, indicating the need to measure the co-expression of p16 and Ki-67 to increase the sensitivity and specificity of the cytological method. Similar reproducibility when comparing the manual and automatic evaluation of the "double label" suggests that the neural network algorithm can currently help in decision support rather than replace the cytologist at the diagnostic stage.

**Keywords:** Pap smear, p16/Ki-67 co-expression, neural network analysis, reproducibility, Bethesda system

**For citation:** Tregubova AV, Tevrukova NS, Ezhova LS, Shamarakova MV, Badlaeva AS, Dobrovolskaya DA, Bayramova GR, Nazarova NM, Shilyaev AY, Asaturova AV. Reproducibility of cytological diagnoses in evaluating liquid cervical smears and immunocytochemical co-expression of p16/Ki-67 using manual and automatic methods. *Gynecology*. 2022;24(6):499–505. DOI: 10.26442/20795696.2022.6.202009

## Введение

Цитологическое исследование мазков шейки матки с окраской по Папаниколу является основополагающим методом скрининга рака шейки матки в Российской Федерации (в соответствии с Приказом Минздрава России от 20.10.2020 №1130н). Данный метод, хотя и не является самым эффективным, все же продемонстрировал значительные результаты в развивающихся странах и позволил значительно снизить заболеваемость раком шейки матки за последние 40 лет [1, 2]. В то же время ПАП-тест подвергается критике в связи с довольно низкой чувствительностью и воспроизводимостью диагнозов между цитологами, даже при наличии значительного практического опыта [3–5]. Следует отметить, что в последние годы все чаще в практику клинического цитолога включаются дополнительные методы, позволяющие улучшить точность диагностики интраэпителиальной патологии шейки матки. Прежде всего речь идет о иммуноцитохимическом исследовании коэкспрессии p16 и Ki-67, доступном в виде коммерческих наборов. В ряде исследований показано, что воспроизводимость

диагнозов с применением иммуноцитохимических исследований повышается [6, 7]. Также в последнее десятилетие разработано множество систем автоматического анализа цитологических мазков для поддержки принятия решений. Применение таких систем призвано повысить воспроизводимость цитологических диагнозов, и в некоторых исследованиях удалось показать положительные результаты [8–11].

Несмотря на успех первичного скрининга на вирус папилломы человека (ВПЧ) во многих развитых странах и перечисленные недостатки, цитологический скрининг продолжает оставаться ведущим методом скрининга в России, однако исследований воспроизводимости цитологических диагнозов в российской популяции практически не проводилось [12]. А воспроизводимость диагнозов с применением «двойной метки» и автоматической оценки в России не исследовалась совсем. В связи с этим решено провести исследование воспроизводимости цитологических диагнозов, поставленных несколькими цитологами при изучении мазков, окрашенных по Папаниколу, воспроизводимости оценки двойного окрашивания на p16 и Ki-67, а также ре-

**Байрамова Гюльдана Рауфовна** – д-р мед. наук, зав. по клин. работе научно-поликлинического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0003-4826-661X

**Назарова Нисо Мирзоевна** – д-р мед. наук, вед. науч. сотр. научно-поликлинического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0001-9499-7654

**Шильяев Алексей Юрьевич** – канд. мед. наук, врач – акушер-гинеколог научно-поликлинического отд-ния ФГБУ «НМИЦ АГП им. акад. В.И. Кулакова». ORCID: 0000-0001-7200-2708

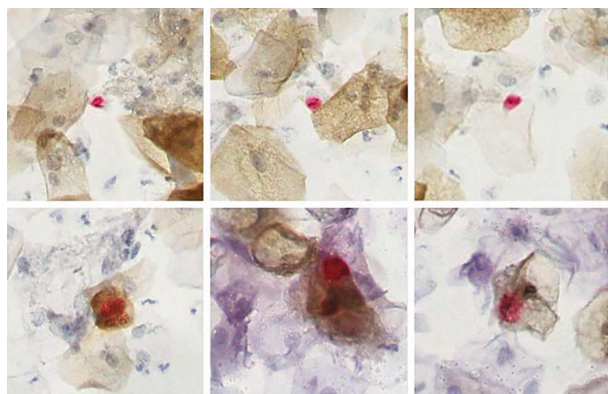
**Guldana R. Bayramova** – D. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0003-4826-661X

**Niso M. Nazarova** – D. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0001-9499-7654

**Alexey Yu. Shilyaev** – Cand. Sci. (Med), Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology. ORCID: 0000-0001-7200-2708

**Рис. 1. Примеры автоматической оценки экспрессии p16 и Ki-67.** Верхний ряд: ложноположительная интерпретация: есть клетки с позитивной экспрессией Ki-67 (красное окрашивание) и позитивной экспрессией p16 (коричневое окрашивание), но совместной экспрессии маркеров нет. Нижний ряд: верная интерпретация, есть клетки с одновременным красным и коричневым окрашиванием, положительная «двойная метка».

**Fig. 1. Examples of automatic evaluation of p16 and Ki-67 expression.** Top row: false positive interpretation; there are cells with positive expression of Ki-67 (red staining) and positive expression of p16 (brown staining) but no co-expression of markers. Bottom row: correct interpretation, there are cells with simultaneous red and brown staining, positive "double label."



зультатов, полученных при применении автоматической системы оценки шеечных мазков.

## Материалы и методы

Проведенное исследование – поперечное, исследовались цитологические мазки, окрашенные по Папаниколау, приготовленные методом жидкостной цитологии по системе SurePath (Becton Dickinson) в 1-м патологоанатомическом отделении ФГБУ «НМИЦ АГП им. В.И. Кулакова» в период с 1 января 2019 по 1 января 2022 г. Для оценки воспроизводимости отобраны 100 мазков (50 пар ПАП-мазок–иммуноцитохимический препарат), которые просматривались 5 цитологами с опытом работы в области гинекологической клинической цитологии не менее 3 лет. Цитологические заключения давались по системе Бетесда (2014 г.), а также с выделением категорий норма/патология. В качестве иммуноцитохимического исследования экспрессии p16 и Ki-67 применялся коммерческий набор CINtec (Roche, США), иммуностейнер Ventana BenchMark Ultra, а для оценки автоматического определения «двойной метки» (коэкспрессии p16 и Ki-67) использовался дополнительный модуль Vision Cyto Pap ICC (Вест Медика, Россия). Коэкспрессию p16 и Ki-67 считали положительной при наличии хотя бы одной клетки с красным окрашиванием ядра и коричневым окрашиванием ядра/цитоплазмы. Данный порог в 1 позитивную клетку применялся как для ручного, так и для автоматического метода оценки иммуноцитохимических препаратов. При неверном определении коэкспрессии нейросетью положительный результат не выбраковывался вне зависимости от количества найденных позитивных клеток (рис. 1).

В качестве статистического критерия использовалась каппа Коэна при попарных сравнениях диагнозов между цитологами или между цитологом и нейросетью и каппа Флейса при оценке воспроизводимости диагнозов между всеми цитологами. Оценивалась воспроизводимость диагнозов по

**Таблица 1. Воспроизводимость цитологических диагнозов по системе Бетесда (2014 г.)**

**Table 1. Reproducibility of cytological diagnoses according to Bethesda system (2014)**

	Цитолог 2	Цитолог 3	Цитолог 4	Цитолог 5
Цитолог 1	0,216 (<0,001)	0,265 (<0,001)	*-0,003 (0,946)	0,055 (0,126)
Цитолог 2		0,274 (<0,001)	0,163 (0,009)	0,05 (0,347)
Цитолог 3			0,098 (0,163)	0,048 (0,026)
Цитолог 4				0,017 (0,596)
Каппа Флейса между всеми цитологами			0,103	$p < 0,001$

Здесь и в табл. 2: \*маркер очень низкого согласия между исследователями.

**Таблица 2. Воспроизводимость цитологических диагнозов при выделении групп нормы и патологии (бинарной оценке)**

**Table 2. Reproducibility of cytological diagnoses using normal/abnormal categories (binary assessment)**

	Цитолог 2	Цитолог 3	Цитолог 4	Цитолог 5
Цитолог 1	0,244 (0,004)	0,377 (<0,001)	0,058 (0,518)	0,124 (0,026)
Цитолог 2		0,354 (<0,001)	0,301 (0,003)	0,124 (0,150)
Цитолог 3			0,188 (0,05)	0,187 (0,002)
Цитолог 4				*-0,042 (0,597)
Каппа Флейса между всеми цитологами			0,174	$p < 0,001$

категориям NILM, ASC-US, LSIL, HSIL, а также по категориям норма/патология (для цитологических мазков, окрашенных по Папаниколау); и по категориям «двойная метка» позитивная/«двойная метка» негативная (для иммуноцитохимических препаратов). Каждый из цитологов оценивал наборы цитологических и иммуноцитохимических препаратов независимо друг от друга. Для всех препаратов, в которых установлен диагноз HSIL, доступны гистологические препараты, подтверждающие диагноз HSIL (диагноз ставился опытным гинекологическим патологом). Оценка уровня воспроизводимости проводилась по следующей шкале: 0,00–0,20 – низкая; 0,21–0,40 – удовлетворительная; 0,41–0,60 – средняя; 0,61–0,80 – хорошая; 0,81–1,00 – отличная [13]. Статистическая обработка полученных результатов проводилась с помощью пакета программного обеспечения SPSS версии 26.0.0.0.

## Результаты

При оценке воспроизводимости четырех категорий цитологических диагнозов по системе Бетесда выявлено, что уровень коэффициента каппа Коэна составил 0,048–0,265, т.е. находился на низком уровне. Общая каппа Флейса между всеми цитологами составила 0,103 (низкий уровень воспроизводимости); табл. 1.

При выделении только двух категорий (норма и патология) уровень воспроизводимости составил от 0,058 до 0,377, т.е. от низкой до удовлетворительной (табл. 2).

При определении «двойной метки» выявлены следующие результаты: при оценке коэкспрессии p16 и Ki-67 выявлена воспроизводимость по каппе Коэна от 0,196 до 0,574, т.е. от низкой до средней, при этом общая каппа Флейса составила 0,407 (средняя); табл. 3.

При сопоставлении результатов оценки каждым из цитологов с нейросетью воспроизводимость по каппе Коэна составила от 0,103 до 0,436 (от низкой до средней); табл. 4.

## Обсуждение

Цитологическое исследование ПАП-мазка остается одной из самых распространенных стратегий скрининга рака шейки матки в мире, особенно в странах с невысоким

экономическим уровнем [14–17]. Известно, что от момента заражения ВПЧ до развития тяжелого поражения шейки матки (HSIL) проходит в среднем 8–10 лет. Таким образом, даже в случае пропуска патологии при первом и даже при втором раунде скрининга все еще остается значительная возможность детекции поражения шейки матки на ранних стадиях [18]. Следует отметить, что традиционный метод приготовления мазка является очень бюджетным исследованием, однако именно жидкостный метод позволяет достичь высокого охвата скрининговым обследованием даже при небольшом количестве высококвалифицированных клинических цитологов, подготовка которых требует значительного количества времени и ресурсов. Показано, что переход на жидкостную технологию приготовления цитологических мазков позволил достичь снижения неопределенных результатов (ASC), увеличения чувствительности и специфичности цитологического метода, а также дал возможность выполнять дополнительные исследования из того же материала, фиксированного в транспортной жидкости виалы (иммуноцитохимическое определение p16 и Ki-67, ВПЧ-типирование) [19–21]. Поэтому, несмотря на высокую стоимость оборудования для жидкостной цитологии, более высокую стоимость расходных материалов, необходимость дополнительного обучения цитологов и цитотехнологов, многие лаборатории успешно внедрили данную методику в свой диагностический процесс, в том числе в России [22].

Наиболее распространенными технологиями для производства жидкостной цитологии в мире, а затем и в России стали ThinPrep (Hologic), SurePath (Becton Dickinson) и CellPrep (Byodyne). Технологии ThinPrep и CellPrep основаны на методе прямой фильтрации, в то время как технология SurePath предполагает предварительное центрифугирование для достижения более «чистых» мазков, лишенных примесей и избыточного количества воспалительных клеток и форменных элементов крови. В настоящем исследовании применялась именно технология SurePath. Следует отметить, что данная технология предполагает выраженное «закругление» клеточных комплексов и краев клеток, что делает их менее похожими на клетки, характерные для традиционных мазков, чем такие технологии, как ThinPrep и CellPrep [23]. В связи с этим клинические цитологи, которые проводят диагностику патологии шейки матки с применением мазков SurePath, должны пройти специальное обучение. Все пять цитологов, принимавших участие в данном исследовании, имеют соответствующие сертификаты, свидетельствующие о том, что они прошли такое обучение.

Тем не менее воспроизводимость диагнозов, поставленных на основе ПАП-мазков в нашем исследовании, являлась довольно низкой и не превышала удовлетворительных значений при сопоставлении заключений отдельных пар цитологов. Следует отметить, что основными факторами, влияющими на невысокий уровень воспроизводимости цитологических мазков, являются количество выделяемых категорий и количество патологических случаев, включенных в исследование [5]. Так, например, в исследовании M. Stoler и соавт., в которое включено относительно небольшое количество мазков с патологическими изменениями, уровень воспроизводимости диагнозов оказался удовлетворительным ( $\kappa=0,46$ ) [3], схожий уровень воспроизводимости показан и в исследовании M. Confortini и соавт. ( $\kappa=0,46$ ) [24]. При большом количестве мазков категории ASC-US уровень воспроизводимости неизменно снижается: так, в исследовании S. Sriamporn и соавт. с большим количеством патологических мазков согласованность диагнозов оказалась умеренной ( $\kappa=0,37$ ) [25].

**Таблица 3. Воспроизводимость оценки коэкспрессии p16 и Ki-67**  
**Table 3. Reproducibility of p16 and Ki-67 co-expression evaluation**

	Цитолог 2	Цитолог 3	Цитолог 4	Цитолог 5
Цитолог 1	0,345 (<0,001)	0,432 (<0,001)	0,520 (<0,001)	0,525 (<0,001)
Цитолог 2		0,574 (<0,001)	0,196 (0,003)	0,497 (<0,001)
Цитолог 3			0,325 (<0,001)	0,567 (<0,001)
Цитолог 4				0,351 (<0,001)
Каппа Флейса между всеми цитологами			0,407	$p<0,001$

**Таблица 4. Воспроизводимость оценки коэкспрессии p16 и Ki-67 между цитологами и нейросетью**  
**Table 4. Reproducibility of p16 and Ki-67 co-expression evaluation between cytologists and neural network**

	Цитолог 1	Цитолог 2	Цитолог 3	Цитолог 4	Цитолог 5
Нейросеть	0,334 (<0,001)	0,103 (<0,046)	0,248 (<0,001)	0,436 (<0,001)	0,295 (<0,001)

В нашем исследовании патологических мазков более 70%, в связи с чем мы полагаем, что невысокие показатели воспроизводимости связаны именно с этим фактором, а не с количеством категорий, выделяемых для исследования, поскольку деление исследуемого материала на две группы (норма и патология) вместо четырех (NILM, ASC-US, LSIL и HSIL) позволило повысить показатель воспроизводимости всего на 0,073, при этом она все равно оставалась низкой. К такому же выводу пришли и H. Hwang и соавт., которые выявили разницу в воспроизводимости цитологических заключений, сделанных с использованием трех и пяти категорий, равную всего лишь 0,04 [5].

Одним из самых распространенных методов повышения воспроизводимости цитологических диагнозов при патологии шейки матки является применение иммуноцитохимического исследования с использованием определения коэкспрессии p16 и Ki-67 [26–28].

Данный тест даже предложен в качестве сортировочного (triage) теста при первичном ВПЧ-скрининге в связи с высокой эффективностью в сочетании с экономической обоснованностью [29]. Оценка воспроизводимости диагнозов при применении иммуноцитохимического теста продемонстрировала более высокие результаты по сравнению с ПАП-тестом. Так, в исследовании M. McMenamin и соавт. коэффициент воспроизводимости каппа составил 0,65–0,91 [30]. В исследовании N. Wentzensen и соавт. коэффициент воспроизводимости к варьировался от 0,65 до 0,81, причем для всех исследователей он составил 0,71, а для цитотехнологов – 0,73 [31]. В исследовании S. Goh и соавт. воспроизводимость результата оценки коэкспрессии p16/Ki-67 оказалась отличной для всех групп наблюдений (0,834–0,977) [32].

Также показано, что с увеличением количества просмотренных иммуноцитохимических препаратов воспроизводимость диагнозов существенно улучшается: от средней ( $\kappa=0,47$ ) при просмотре менее 30 препаратов до отличной ( $\kappa=0,88$ ) при просмотре более 300 препаратов [33]. В некоторых исследованиях, особенно охватывающих несколько лабораторий, однако, отмечалась невысокая воспроизводимость оценки двойной метки, которая оказалась еще ниже для мазков с диагнозами CIN 2+ (0,558 и 0,487 соответственно) [24].

Кроме того, существенно различалась и воспроизводимость диагнозов в отношении различных диагностических категорий: в исследовании M. Venevolò и соавт. коэффици-

ент каппа составил 0,692 для позитивной и 0,641 для негативной категорий мазков (хорошая воспроизводимость), в то время как для неопределенной категории данный показатель стал очень низким ( $\kappa=0,058$ ) [34, 35]. В нашем исследовании мы получили существенное (почти в 3 раза) увеличение показателя воспроизводимости каппа при использовании «двойной метки» по сравнению с воспроизводимостью диагнозов с применением ПАП-мазков. Однако она все еще оставалась на среднем уровне (общая каппа Флейса составила 0,407). Эти результаты свидетельствуют о том, что оценка экспрессии p16/Ki-67 сопряжена с определенными трудностями, что, в частности, касается интерпретации двойного окрашивания в группах клеток и в клетках с небольшим количеством цитоплазмы, а также клеток с бледной цитоплазматической окраской на p16 [30, 35].

К сожалению, преодолеть эти ограничения не удалось и с применением автоматической системы оценки экспрессии p16/Ki-67, которая проводилась с использованием специального модуля программы Vision Cyto Pap ICC. Более того, показатели воспроизводимости каппы Коэна оказались ниже в парах цитолог–нейросеть (0,103–0,436), чем в парах между цитологами (0,196–0,574). Это связано с несколькими факторами, которые в совокупности могут приводить к невысокой воспроизводимости результатов, основанных только на автоматической оценке. В частности, следует отметить, что в большинстве случаев при разработке нейросетевых подходов для автоматического анализа цифровых копий препаратов на этапе рутинной разметки клеточных элементов и окрашенных тканевыми маркерами структур используется труд недостаточно квалифицированных специалистов (студентов или ординаторов), поскольку считается, что этот этап очень трудоемкий и в то же время не требует внимательного анализа каждой клетки/структуры. Однако это не соответствует действительности: на самом деле именно от того, насколько точно и грамотно будет проведена разметка, зависит конечная эффективность работы нейросетевого анализа [36, 37]. Поэтому этот этап требует значительных ресурсов, как финансовых, так и человеческих.

Кроме того, следует отметить, что эволюционно автоматизация процессов в цитологической диагностике шла по пути замены самых простых, рутинных процессов, требующих значительного количества времени, при этом квалификация для их выполнения требовалась минимальная (например, ручное окрашивание препаратов представляет собой последовательное погружение препарата в подготовленные растворы). В связи с этим автоматизация таких процессов проводилась успешно, практически во всех лабораториях в настоящее время в рабочем процессе используются автоматические стейнеры для различных видов окраски цитологических мазков, а также иммуностейнеры для проведения иммуноцитохимических исследований. Поэтому кажется логичным предлагать именно замену самых легких этапов диагностики с помощью нейросети, например узнавание неизменных клеток, а не стараться правильно верифицировать диагностические категории.

С помощью такого же подхода разрабатывались системы для автоматической оценки цитологических мазков компаниями-производителями систем для приготовления жидкостных мазков – модуль ThinPrep Imaging System (Hologic Inc) [38], FocalPoint (Becton Dickinson) [39], Vision Cyto Pap (Вест Медика), применение которых преимущественно основано на исключении из оценки цитологом неизменных клеток с выделением (отбором) подозрительных клеточных элементов и групп клеток. Данные манипуляции существенно облегчают работу цитолога, позволяя ему не про-

сматривать все поля зрения, а сконцентрироваться только на тех клетках, которые не совпадают с неизменными по каким-либо характеристикам. Однако и FocalPoint, и Vision Cyto Pap могут предложить цитологу для оценки очень большое количество патологически измененных клеток, не сокращая, таким образом, время просмотра препарата. В то же время ThinPrep Imaging System предлагает оценить только 22 клетки/группы клеток, причем к необходимым для просмотра полям зрения будут относиться и железистые клетки в комплексах. Таким образом, остается вероятность для цитолога не увидеть наиболее измененные клетки и записать диагноз.

В какой-то степени данные ограничения касаются и автоматической оценки иммуноцитохимических исследований. Подразумевается, что цитолог должен оценить все клетки, которые отобраны алгоритмом нейросети как подозрительные в отношении наличия коэкспрессии p16/Ki-67. В нашем исследовании мы принимали за положительные результаты те образцы, где хотя бы одна клетка расценена автоматическим модулем как положительная в отношении «двойной метки». Именно тот факт, что на данном этапе не происходил этап «выбраковки» ложноположительных результатов, мог привести к невысокой согласованности цитологов и нейросети при определении экспрессии иммуноцитохимических маркеров.

Следует отметить, что в мировой литературе пока еще не велико количество публикаций, посвященных сравнению ручного и автоматического методов оценки коэкспрессии p16 и Ki-67, несмотря на то что Управление по контролю пищевых продуктов и лекарств в США одобрило коммерческий тест CINtec для определения данных маркеров в качестве скринингового метода [33]. В тех исследованиях, которые опубликованы, результаты касаются не только мазков шейки матки, но и анальной цитологии. В них отмечаются следующие отличия: во-первых, в качестве порогового значения в данных исследованиях принималось наличие нескольких (от 3 до 5) клеток с коэкспрессией p16/Ki-67 (в то время как мы в нашем исследовании определили одинаковый порог в 1 позитивную клетку как для ручного, так и для автоматического этапов). Во-вторых, авторы не приводят в своих результатах данных о воспроизводимости диагнозов конкретных пациентов, а лишь указывают на общее количество положительных или отрицательных в отношении «двойной метки» мазков [7]. Таким образом, можно предположить, что в настоящее время место автоматических систем анализа цитологических мазков – преимущественно в области разработки системы поддержки принятия решений и выделения областей интереса, а не в замене цитолога на этапе классификации диагностических категорий или оценки экспрессии иммуноцитохимических маркеров.

## Заключение

В настоящем исследовании воспроизводимость цитологических заключений с применением системы Бетесда 2014 г., а также двух категорий (норма/патология) на основании исследования ПАП-мазков, приготовленных методом жидкостной цитологии (SurePath), оказалась низкой. С нашей точки зрения, такие результаты прежде всего связаны с большим количеством патологических мазков, включенных в исследование. Применение иммуноцитохимического метода позволило существенно (в 3 раза) увеличить воспроизводимость диагнозов, что свидетельствует о необходимости использования определения коэкспрессии p16 и Ki-67 для повышения чувствительности и специфичности цитологического метода. Сравнение ручной и автоматической оцен-

ки «двойной метки» выявило сопоставимые результаты при сравнении воспроизводимости между цитологами и между цитологами и нейросетевым алгоритмом. Это свидетельствует о том, что в настоящее время данный алгоритм может помочь в основном в области поддержки принятия решения, а не заменить цитолога на диагностическом этапе.

**Раскрытие интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

**Disclosure of interest.** The authors declare that they have no competing interests.

**Вклад авторов.** Авторы декларируют соответствие своего авторства международным критериям ICMJE. Все авторы в равной степени участвовали в подготовке публикации: разработка концепции статьи, получение и анализ фактических данных, написание и редактирование текста статьи, проверка и утверждение текста статьи.

**Authors' contribution.** The authors declare the compliance of their authorship according to the international ICMJE criteria. All authors made a substantial contribution to the conception of the work, acquisition, analysis, interpretation of data for the work, drafting and revising the work, final approval of the version to be published and agree to be accountable for all aspects of the work.

**Источник финансирования.** Исследование выполнено при финансовой поддержке РФФИ и Правительства Москвы в рамках научного проекта №21-315-70048.

**Funding source.** The study was performed with the financial support of the Russian Foundation for Basic Research (RFBR) and the Government of Moscow within the scientific project №21-315-70048.

## Литература/References

- Siegel RL, Miller KD, Fuchs HE. Cancer statistics, 2022. *CA Cancer J Clin.* 2022;72(1):7-33.
- Boring CC, Squires TS, Tong T. Cancer statistics, 1999. *CA Cancer J Clin.* 1999;49(1):8-31.
- Stoler MH, Schiffman M. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA.* 2001;285(11):1500-5.
- Selvaggi SM. Implications of low diagnostic reproducibility of cervical cytologic and histologic diagnoses. *JAMA.* 2001;285(11):1506-8.
- Hwang H, Follen M, Guillaud M, et al. Cervical cytology reproducibility and associated clinical and demographic factors. *Diagn Cytopathol.* 2020;48(1):35-42.
- Kloboves Prevodnik V, Jerman T, Nolde N, et al. Interobserver variability and accuracy of p16/Ki-67 dual immunocytochemical staining on conventional cervical smears. *Diagn Cytopathol.* 2019;14(1):48.
- Wentzensen N, Lahrmann B, Clarke MA, et al. Accuracy and Efficiency of Deep-Learning-Based Automation of Dual Stain Cytology in Cervical Cancer Screening. *J Natl Cancer Inst.* 2021;113(1):72-9.
- Dey P. Artificial neural network in diagnostic cytology. *CytoJournal.* 2022;19:146.
- Sanyal P, Barui S, Deb P, Sharma HC. Performance of A Convolutional Neural Network in Screening Liquid Based Cervical Cytology Smears. *J Cytol.* 2019;36(3):146.
- Mohammed MA, Abdurahman F, Ayalew YA. Single-cell conventional pap smear image classification using pre-trained deep neural network architectures. *BMC Biomed Eng.* 2021;3(1):1-8.
- Zhang L, Lu L, Member S, et al. DeepPap: Deep Convolutional Networks for Cervical Cell Classification. *IEEE J Biomed Health Inform.* 2017;21(6):1633-43.
- Фириченко С.В., Манухин И.Б., Роговская С.И., Манухина Е.И. «Подводные камни» цервикального скрининга. *Доктор.Ру.* 2018;2(146):26-34 [Firichenko SV, Manukhin IB, Rogovskaya SI, Manukhina EI. Pitfalls in Cervical Screening. *Doctor.Ru.* 2018;2(146):26-34 (in Russian)].
- Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics.* 1977;33(1):159.
- Swales AL, Hossler CE, Kesterson JP. Pathway to the Papanicolaou smear: The development of cervical cytology in twentieth-century America and implications in the present day. *Gynecol Oncol.* 2019;154(1):3-7.
- Salehiniya H, Momenimovahed Z, Allahqoli L, et al. Factors related to cervical cancer screening among Asian women. *Riv Eur Sci Med Farmacol.* 2021;25(19):6109-22.
- Cudjoe J, Nkimbeng M, Turkson-Ocran RA, et al. Understanding the Pap Testing Behaviors of African Immigrant Women in Developed Countries: A Systematic Review. *J Immigr Minor Health.* 2021;23(4):840-56.
- Chin SS, Jamonek Jamhuri NAB, Hussin N, et al. Factors influencing pap smear screening uptake among women visiting outpatient clinics in Johor. *Malays Fam Physician.* 2022;17(2):46-55.
- Settakorn J, Rangdaeng S, Preechapornkul N, et al. Interobserver reproducibility with LiquiPrep liquid-based cervical cytology screening in a developing country. *Asian Pac J Cancer Prev.* 2008;9(1):92-6.
- Strander B, Andersson-Ellström A, Milsom I, et al. Liquid-based cytology versus conventional Papanicolaou smear in an organized screening program: a prospective randomized study. *Cancer.* 2007;111(5):285-91.
- Nishio H, Iwata T, Nomura H, et al. Liquid-based cytology versus conventional cytology for detection of uterine cervical lesions: a prospective observational study. *Jpn J Clin Oncol.* 2018;48(6):522-8.
- Sharma P, Gupta P, Gupta N, et al. Evaluation of the Performance of CinTec® PLUS in SurePath™ Liquid-Based Cervico-Vaginal Samples. *Turk Patoloji Dergisi.* 2021;37(1):32-8.
- Сухих Г.Т., Прилепская В.Н., Асатурова А.В., и др. Диагностика, лечение и профилактика цервикальных интраэпителиальных неоплазий. М., 2020 [Sukhikh GT, Prilepskaia VN, Asaturova AV, et al. Diagnostika, lecheniie i profilaktika tservikal'nykh intraepitel'nykh neoplazii. Moscow, 2020 (in Russian)].
- Sharma J, Toi P, Siddaraju N, et al. A comparative analysis of conventional and SurePath liquid-based cervicovaginal cytology: A study of 140 cases. *J Cytol.* 2016;33(2):80-4.
- Confortini M, Bondi A, Cariaggi MP, et al. Interlaboratory reproducibility of liquid-based equivocal cervical cytology within a randomized controlled trial framework. *Diagn Cytopathol.* 2007;35(9):541-4.
- Sriamporn S, Kritpetcharat O, Nieminen P, et al. Kritpetcharat O. Consistency of cytology diagnosis for cervical cancer between two laboratories. *Asian Pac J Cancer Prev.* 2005;6(2):208-12.
- Tjalma WAA. Diagnostic performance of dual-staining cytology for cervical cancer screening: A systematic literature review. *Eur J Obstet Gynecol Reprod Biol.* 2017;210:275-80.
- Bergeron C, Ikenberg H, Sideri M, et al. Prospective evaluation of p16/Ki-67 dual-stained cytology for managing women with abnormal Papanicolaou cytology: PALMS study results. *Cancer Cytopathol.* 2015;123(6):373-81.
- Li Y, Fu Y, Cheng B, et al. A Comparative Study on the Accuracy and Efficacy Between Dalton and CINtec® PLUS p16/Ki-67 Dual Stain in Triaging HPV-Positive Women. *Front Oncol.* 2022;11:815213.
- Han Q, Guo H, Geng L, Wang Y. p16/Ki-67 dual-stained cytology used for triage in cervical cancer opportunistic screening. *Chin J Cancer.* 2020;32(2):208.
- McMenamin M, McKenna M, McDowell A, et al. Intra- and inter-observer reproducibility of CINtec® PLUS in ThinPrep® cytology preparations. *Cytopathology.* 2017;28(4):284-90.
- Wentzensen N, Fetterman B, Tokugawa D, et al. Interobserver reproducibility and accuracy of p16/Ki-67 dual-stain cytology in cervical cancer screening. *Cancer Cytopathol.* 2014;122(12):914-20.
- Goh ST, TayKah T, Lim L. Inter-observer Variability of CINtec PLUS Dual Staining for p16/ki67. *J Am Soc Cytopathol.* 2017;6(5):S30.
- Hammer A, Gustafson LW, Christensen PN, et al. Implementation of p16/Ki67 dual stain cytology in a Danish routine screening laboratory:

- Importance of adequate training and experience. *Cancer Med.* 2020;9(21):8235-42.
34. Benevolo M, Allia E, Gustinucci D, Montaguti A. Interobserver reproducibility of cytologic p16INK4a /Ki-67 dual immunostaining in human papillomavirus-positive women. *Cancer Cytopathol.* 2017;125(3):212-20.
35. Benevolo M, Mancuso P, Allia E, et al. Interlaboratory concordance of p16/Ki-67 dual-staining interpretation in HPV-positive women in a screening population. *Cancer Cytopathol.* 2020;128(5):323-32.
36. Sornapudi S, Addanki R, Stanley RJ, et al. Automated Cervical Digitized Histology Whole-Slide Image Analysis Toolbox. *J Pathol Informatics.* 2021;12(1):26.
37. Kanavati F, Hirose N, Ishii T, et al. A Deep Learning Model for Cervical Cancer Screening on Liquid-Based Cytology Specimens in Whole Slide Images. *Cancers.* 2022;14(5):1159.
38. Özcan Z, Kimiloğlu E, Akyıldız İğdem A, Erdoğan N. Comparison of the Diagnostic Utility of Manual Screening and the ThinPrep Imaging System in Liquid-Based Cervical Cytology. *Turk Patoloji Dergisi.* 2020;36(2):135-41.
39. Nuttall DS, Hillier S, Clayton HR, et al. A retrospective validation of the FocalPoint GS slide profiler NFR technology by analysis of interval disease outcomes compared with manual cytology. *Cancer Cytopathol.* 2019;127(4):240-6.

---

Статья поступила в редакцию / The article received: 10.11.2022

Статья принята к печати / The article approved for publication: 16.12.2022