

## بررسی میزان توافق بین فردی در تشخیص تغییرات غیرطبیعی سلول‌های پوششی دهانه رحم در نمونه‌های پاپ‌اسمیر

تاریخ دریافت مقاله: ۱۳۸۹/۱۲/۱۵ تاریخ پذیرش: ۱۳۹۰/۰۱/۲۹

### چکیده

نرگس ایزدی مود،<sup>۱\*</sup> سهیلا سرمدی،<sup>۱</sup> فرزانه حیدری‌فرزان،<sup>۲</sup> هایده حائری،<sup>۳</sup> زهرا فروهش تهرانی<sup>۲</sup>

۱- گروه پاتولوژی، بیمارستان زنان

۲- گروه پاتولوژی، بیمارستان شریعتی

۳- گروه پاتولوژی، بیمارستان امام‌خمینی

دانشگاه علوم پزشکی تهران، تهران، ایران.

**زمینه و هدف:** کارسینوم دهانه رحم دومین بدخیمی مسبب مرگ و میر زنان در جهان و شایع‌ترین بدخیمی در کشورهای در حال توسعه است. از آن‌جا که بسیاری از زنان مبتلا کارسینوم سرویکس قبلاً توسط پاپ‌اسمیر غربالگری نشده بودند، پزشکان بر این عقیده‌اند که پاپ‌اسمیر از دقت بالایی برخوردار است، اما مسایلی مانند عدم تکرارپذیری کارایی آن را به چالش کشیده است و تاکنون مطالعات چندانی برای بررسی میزان تکرارپذیری صورت نگرفته است. **روش بررسی:** ۱۶۲ اسلاید پاپ‌اسمیر با تشخیص‌های: ASC-US، ASC-H، LSIL، HSIL، SCC، AGC و آدنوکارسینوما از فایل سیتولوژی بیمارستان زنان انتخاب گردیده و سپس توسط پاتولوژیست مرجع مورد بازخوانی قرار گرفت و در یکی از دسته‌های تشخیصی ذکر شده قرار داده شدند. این ۱۶۲ اسلاید به طور مستقل توسط سه پاتولوژیست دیگر نیز مورد بازبینی قرار گرفتند. داده‌های مطالعه مورد آنالیز آماری قرار گرفتند. **یافته‌ها:** بیشترین و کم‌ترین میزان تکرارپذیری بین فردی به ترتیب در دسته تهاجمی‌ها (SCC) به علاوه آدنوکارسینوم) با  $K=0/53$  و دسته HSIL بود ( $K=0/19$ ). در کل میزان تکرارپذیری در بین مشاهده‌گران اندک بود ( $K=0/26$ ). توافق نسبی تا کامل در ۴۲٪ اسلایدها دیده شد اما در ۱۰/۵٪ اسلایدها هیچ‌گونه توافقی بین چهار مشاهده‌گر وجود نداشت. **نتیجه‌گیری:** اندک بودن میزان توافق بین فردی با برخی مطالعات دیگر از این دست هم‌خوانی داشت. توصیه می‌شود که مطالعاتی با حجم نمونه بالاتر در تمام گروه‌های تشخیصی پاپ‌اسمیر انجام شود تا نتایج قابل استنادتری برای مداخلاتی جهت کاهش اختلاف نظر در تفسیر پاپ‌اسمیر به دست آید.

**کلمات کلیدی:** پاپ‌اسمیر، تکرارپذیری، ضایعات اپی‌تلیالی سلول سنگفرشی، ضایعات اپی‌تلیالی سلول غددی.

\* نویسنده مسئول: تهران، خیابان کریم‌خان زند، خیابان استاد نجات‌اللهی شمالی، بیمارستان زنان، بخش پاتولوژی

تلفن: ۰۲۱-۸۸۹۰۶۷۶۷

email: nizadimood@yahoo.com

### مقدمه

رحم برای تشخیص سرطان قبل از مراحل تهاجمی به کار می‌رود و از آن‌جا که بسیاری از زنانی که به این سرطان مبتلا شده‌اند قبلاً توسط پاپ‌اسمیر غربالگری نشده بودند، پزشکان بر این عقیده‌اند که پاپ‌اسمیر از دقت بالایی در تشخیص ضایعات بدخیم دهانه رحم برخوردار است، اما مسایلی هم‌چون تفسیرهای منفی کاذب، مثبت کاذب و عدم تکرارپذیری، میزان کارایی آن را به چالش کشیده است.<sup>۱</sup> برای یک نظام هماهنگ گزارش‌نویسی اسمیرهای دهانه رحم، تاکنون سیستم‌های طبقه‌بندی مختلفی به وجود آمده است: از جمله طبقه‌بندی پاپانیکولا، طبقه‌بندی WHO و سیستم طبقه‌بندی Bethesda.<sup>۲</sup> جوامع علمی تمایل به این موضوع دارند که سیتوپاتولوژیست‌ها به صورت یک‌دست عمل کرده و برای تفسیر نمونه‌ها از یک راه‌کار

سرطان دهانه رحم (Cancer of cervix) دومین بدخیمی شایع در زنان جهان و شایع‌ترین آن‌ها در کشورهای در حال توسعه است. شیوع این سرطان و مرگ و میر ناشی از آن در کشورهای توسعه یافته به لطف برنامه پیشگیری پاپ‌اسمیر کاهش یافته است،<sup>۱</sup> در حالی که در کشورهای در حال توسعه به دلیل کیفیت پایین استانداردهای غربالگری برای پیشگیری، این سرطان هم‌چنان از میزان مرگ و میر بالایی برخوردار است.<sup>۲</sup> تشخیص سرطان دهانه رحم و مرحله‌بندی آن توسط یک سلسله کارهای تشخیصی از جمله سیتولوژی، کولپوسکوپی، هیستولوژی، یافته‌های بالینی و تصویربرداری انجام می‌پذیرد. تست پاپ‌اسمیر با برداشت سلول‌های غیرطبیعی از دهانه

(AGC). پس از بازبینی، تعداد نمونه‌ها در هر دسته تشخیصی به قرار زیر شد که در گروه ASC-US، ۲۹ ASC-H، ۳۳ LSIL، ۲۲ HSIL، ۱۶ SCC، ۱۲ AGC و ۳۷ آدنوکارسینوم ۱۳ نمونه قرار گرفت. نمونه‌ها به‌طور تصادفی و با استفاده از جدول اعداد تصادفی از یک تا ۱۶۲ شماره‌گذاری گردیدند. کلیه اسلایدها در شرایط مساوی به همراه اطلاعات سن بیماران، در اختیار پاتولوژیست‌ها قرار گرفتند. چهار پاتولوژیست از اعضای هیئت علمی دانشگاه تهران با سطوح مختلف تجربه (۱۷-۶ سال) در این مطالعه شرکت کردند که برای تشخیص لام‌ها با هم ارتباط نداشته و به‌طور جداگانه به بررسی اسلایدها پرداختند. در این مطالعه از جزوات آموزشی استفاده نشد و پاتولوژیست‌ها بر اساس زمینه ذهنی خود بر مبنای معیارهای سیستم بتسدا که به طور روزانه در آزمایشگاه خود به کار می‌گرفتند به دسته‌بندی ضایعات دهانه رحم پرداختند. داده‌های این مطالعه بر اساس آنالیز آماری Kappa تجزیه و تحلیل شدند. آنالیز توسط دو نرم‌افزار SPSS ویراست ۱۶ و STATA ویراست ۸ انجام شد. برای آنالیز توافق تشخیصی از آماره کاپا استفاده شد. ارزش‌های مختلف آماره کاپا به شرح زیر ارزیابی می‌گردند.

۰/۲-۰ ضعیف، ۰/۴-۰/۲۱ اندک، ۰/۶-۰/۴۱ به نسبت خوب، ۰/۸-۰/۶۱ خوب، ۰/۸۱-۰/۹۲ خیلی خوب و بیش از ۰/۹۲ عالی. ارزش کاپا از ۱- تا یک متغیر است. آماره کاپا آماره‌ای است که برای ارزیابی میزان توافق بین دو یا بیش از دو مشاهده‌گر به کار می‌رود. برای استفاده از این آماره، نویسندگان مختلف راهنماهای متفاوتی را عنوان کرده‌اند، ارزیابی نتایج ما براساس راهنمای دوسون و Trapp<sup>۱۶</sup>، صورت گرفت (راهنمای مشابهی توسط Koch و Landis<sup>۱۷</sup> نیز مطرح شده است).

براساس راهنمای دوسون، نتایج منفی وقتی حاصل می‌شود که میزان توافق حتی از میزان مورد انتظار کاملاً تصادفی نیز کمتر باشد و ارزش صفر برای کاپا نشان‌دهنده توافق کاملاً تصادفی است. در تحلیل آماری  $P < 0/05$  دلیل بر اختلاف معنی‌دار ارزش کاپا با عدد صفر و در غیر این صورت کاپا بی‌ارزش تلقی می‌شود.

### یافته‌ها

برای ارزیابی توافق بین فردی علاوه بر مقایسه نتایج دو به دو بین مشاهده‌گران میزان توافق به صورت کلی نیز در برنامه STATA

تشخیصی واحد استفاده کرده تا گزارشات آن‌ها با تکرارپذیری بیشتری همراه باشند. با وجود پیروی از چنین سیستم‌های واحد طبقه‌بندی میزان توافقات درون‌فردی و بین‌فردی چه در طبقه‌بندی ضایعات و چه در تعیین میزان کفایت نمونه در تست پاپ‌اسمیر همواره بحث‌برانگیز بوده است و مطالعات متعدد حاکی از درجات مختلف و عمدتاً پایینی از این توافقات می‌باشند. طبق تحقیقات انجام شده در سال‌های گذشته، تکرارپذیری در میان مشاهده‌گرهایی که حتی در سایر زمینه‌ها از مهارت خوبی برخوردار بوده‌اند، کم گزارش شده است.<sup>۴-۱۵</sup> بعضی مطالعات صورت گرفته در سال‌های اخیر نیز حاکی از همین مسئله بوده است اما در کل به نظر می‌رسد که مطالعه کافی در این زمینه مهم صورت نپذیرفته و اطلاعات اندکی در مورد تکرارپذیری در پاپ‌اسمیر در دسترس است<sup>۱۶</sup> که این خود به تنهایی پرداختن به این مقوله را امری ضروری می‌نماید. لذا هدف این پژوهش آن است که میزان توافق در طبقات مختلف تشخیصی را مورد بررسی قرار داده و بدین ترتیب، به طور تقریباً کم سابقه‌ای نتایج مستقلی در توافق در تمام گروه‌های تشخیصی به دست دهد.

### روش بررسی

این مطالعه از نوع توصیفی-تحلیلی می‌باشد، و به صورت مقطعی انجام گرفته است. با مراجعه به دفاتر سیتولوژی بخش پاتولوژی بیمارستان زنان تهران نمونه‌های پاپ‌اسمیر مربوط به بیماران مراجعه کننده به این بیمارستان از سال ۱۳۸۳ تا سال ۱۳۸۸ با تشخیص ضایعات مختلف ایتراپی تلیال دهانه رحم انتخاب شدند. سپس توسط یکی از پاتولوژیست‌های با تجربه‌تر به عنوان پاتولوژیست مرجع مورد بازخوانی و در یکی از گروه‌های تشخیصی سیتولوژی بر اساس طبقه‌بندی بتسدا قرار گرفته و کدگذاری گردیدند. به این ترتیب تعداد ۱۶۲ اسلاید به عنوان حجم نمونه انتخاب شد که در مقایسه با دیگر مطالعات صورت گرفته در این زمینه، قابل قبول می‌باشد. علاوه بر ارزیابی توافق تشخیصی در هفت گروه تشخیصی ذکر شده بر اساس بتسدا، جهت افزایش میزان تکرارپذیری، تعداد دسته‌ها به این گروه‌ها تقلیل داده شد: گروه اسکواموس (شامل: ASC-US، ASC-H، LSIL، HSIL و SCC) در مقابل گروه غددی (شامل: AGC و آدنوکارسینوم)، گروه تهاجمی (شامل: SCC و آدنوکارسینوم) در مقابل غیرتهاجمی (شامل: ASC-US، ASC-H، LSIL، HSIL و

بین تمام اسلایدها، در ۱۳ مورد (۸٪ از کل نمونه‌ها)، مشاهده‌گران هیچ‌گونه توافقی با هم نداشتند (عدم توافق کامل) و هر یک از مشاهده‌گران یک تشخیص متفاوت را برای اسلاید مورد نظر مطرح کرده بودند که نمونه‌هایی از آن‌ها در شکل‌های ۱ و ۲ نشان داده شده است. در ۵۵ مورد (۳۴٪ از اسلایدها) سه نفر از چهار نفر روی یک تشخیص واحد اتفاق نظر داشتند (توافق نسبی) که یکی از این

محاسبه شد. جداول ۱ تا ۳ به ترتیب نشانگر توافق تشخیصی بین مشاهده‌گر مرجع (مشاهده‌گر اول) با مشاهده‌گر دوم و سوم و چهارم می‌باشد که ابتدا بر اساس هفت گروه تشخیصی انجام و سپس گروه‌های تشخیصی به دو گروه اسکواموس و غددی و نیز مهاجم و غیر مهاجم تقسیم شد و نتایج مقایسه گردید. جدول ۴ نشانگر توافق کلی بین همه مشاهده‌گران (محاسبه شده با برنامه Stata) می‌باشد. از

جدول-۱: توافق تشخیصی بین مشاهده‌گر مرجع با مشاهده‌گر دوم

تشنیص	ارزش کاپا	توافق تشنیصی	خطای استاندارد	Approx.sig
ASC-US	۰/۴۶	نسبتاً خوب	۰/۰۹	<۰/۰۰۰۱
ASC-H	۰/۳۳	اندک	۰/۰۹	<۰/۰۰۰۱
LSIL	۰/۷۱	خوب	۰/۰۷	<۰/۰۰۰۱
HSIL	۰/۲۰	ضعیف	۰/۰۸	۰/۰۰۴
SCC	۰/۶۵	خوب	۰/۱۰	<۰/۰۰۰۱
AGC	۰/۲۳	اندک	۰/۰۸	۰/۰۰۱
آدنوکارسینوم	۰/۳۷	اندک	۰/۱۳	<۰/۰۰۰۱
اسکواموس	۰/۳۶	اندک	۰/۰۷	<۰/۰۰۰۱
غددی	۰/۴۴	نسبتاً خوب	۰/۰۷	<۰/۰۰۰۱
تهاجمی	۰/۶۹	خوب	۰/۰۷	<۰/۰۰۰۱
غیر تهاجمی	۰/۶۵	خوب	۰/۰۷	<۰/۰۰۰۱

ASC-US= Atypical Squamous Cells of Undetermined Significance, ASC-H= Atypical Squamous Cells, cannot exclude High-grade squamous intraepithelial lesion, AGC= Atypical Glandular Cells, LSIL= Low-grade Squamous Intraepithelial Lesion, HSIL= High-grade Squamous Intraepithelial Lesion, SCC= Squamous Cell Carcinoma

جدول-۲: توافق تشخیصی بین مشاهده‌گر مرجع با مشاهده‌گر سوم

تشنیص	ارزش کاپا	توافق تشنیصی	خطای استاندارد	Approx.sig
ASC-US	۰/۴۷	غیر قابل ارزیابی	۰/۰۷	۰/۵۲
ASC-H*	-	-	-	-
LSIL	۰/۱۷	ضعیف	۰/۰۹	۰/۰۲
HSIL	۰/۱۹	ضعیف	۰/۱۱	۰/۰۱۴
SCC	۰/۲۲	اندک	۰/۱۴	۰/۰۰۱
AGC	۰/۳۱	اندک	۰/۰۸	۰/۰۰۰
آدنوکارسینوم	۰/۶۴	خوب	۰/۱۲	<۰/۰۰۰۱
اسکواموس	۰/۲۸	اندک	۰/۰۷	<۰/۰۰۰۱
غددی	۰/۴۶	نسبتاً خوب	۰/۰۷	<۰/۰۰۰۱
تهاجمی	۰/۴۲	نسبتاً خوب	۰/۱۰	<۰/۰۰۰۱
غیر تهاجمی	۰/۱۱	غیر قابل ارزیابی	۰/۰۷	۰/۱۲۲

\* از آنجا که مشاهده‌گر سوم هیچ یک از نمونه‌ها را در دسته ASC-H قرار نداده بودند میزان توافق در این دسته تشخیصی برای ایشان غیرممکن بود.

جدول- ۳: توافق تشخیصی بین مشاهده گر مرجع با مشاهده گر چهارم

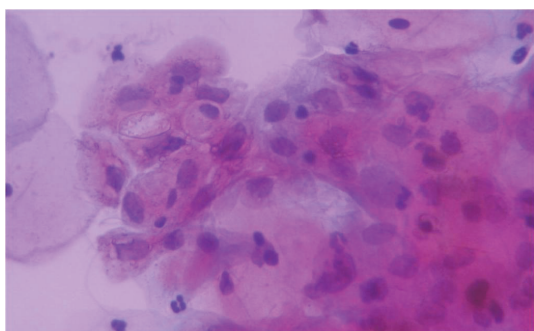
تشیخص	ارزش کاپا	توافق تشخیصی	خطای استاندارد	Approx.sig
ASC-US	۰/۳۸	اندک	۰/۰۹	<۰/۰۰۰۱
ASC-H	۰/۳۲	اندک	۰/۰۹	<۰/۰۰۰۱
LSIL	۰/۶۸	خوب	۰/۰۸	<۰/۰۰۰۱
HSIL	۰/۳۳	اندک	۰/۱۰	<۰/۰۰۰۱
SCC	۰/۵۸	نسبتاً خوب	۰/۱۰	<۰/۰۰۰۱
AGC	۰/۲۲	اندک	۰/۰۸	۰/۰۰۴
آدنوکارسینوم	۰/۵۲	نسبتاً خوب	۰/۱۲	<۰/۰۰۰۱
اسکواموس	۰/۳۷	اندک	۰/۰۷	<۰/۰۰۰۱
غددی	۰/۴۰	اندک	۰/۰۷	<۰/۰۰۰۱
تهاجمی	۰/۶۳	خوب	۰/۰۷	<۰/۰۰۰۱
غیر تهاجمی	۰/۶۲	خوب	۰/۰۷	<۰/۰۰۰۱

آزمون آماری کاپا،  $P < ۰/۰۵$  معنی دار می باشد.

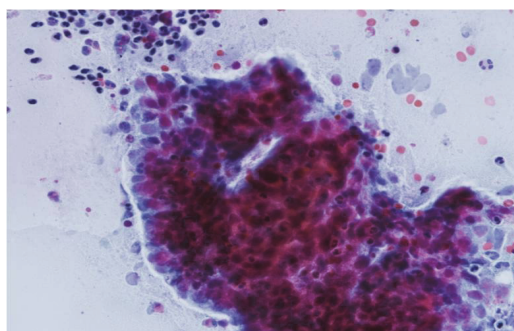
جدول- ۴: توافق کلی بین همه مشاهده گران

تشیخص	ارزش کاپا	توافق تشخیصی	Z	Approx.sig
ASC-US	۰/۲۶	اندک	۸/۱۱	<۰/۰۰۰۱
ASC-H	۰/۱۹	ضعیف	۶/۱۰	<۰/۰۰۰۱
LSIL	۰/۴۳	نسبتاً خوب	۱۳/۵۲	<۰/۰۰۰۱
HSIL	۰/۱۷	ضعیف	۵/۳۹	<۰/۰۰۰۱
SCC	۰/۳۸	اندک	۱۱/۹۹	<۰/۰۰۰۱
AGC	۰/۲۳	اندک	۷/۳۰	<۰/۰۰۰۱
آدنوکارسینوم	۰/۴۱	نسبتاً خوب	۱۳/۰۶	<۰/۰۰۰۱
اسکواموس	۰/۲۷	اندک	۸/۴۷	<۰/۰۰۰۱
غددی	۰/۳۷	اندک	۱۱/۶۷	<۰/۰۰۰۱
تهاجمی	۰/۵۳	نسبتاً خوب	۱۶/۶۶	<۰/۰۰۰۱
غیر تهاجمی	۰/۳۲	اندک	۱۰/۰۵	<۰/۰۰۰۱

از بین تمام اسلایدها، در ۱۳ مورد (۸/۸ از کل نمونه‌ها)، مشاهده گران هیچ گونه توافقی با هم نداشتند (عدم توافق کامل) و هر یک از مشاهده گران یک تشخیص متفاوت را برای اسلاید مورد نظر مطرح کرده بودند که نمونه‌هایی از آن‌ها در شکل‌های ۱ و ۲ نشان داده شده است

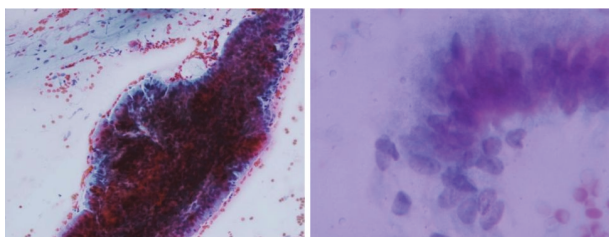


شکل- ۲: عدم توافق کامل، مشاهده گر ۱: ASC-US، مشاهده گر ۲: LSIL، مشاهده گر ۳: بدون تغییرات اپی تلیالی، مشاهده گر ۴: AGC

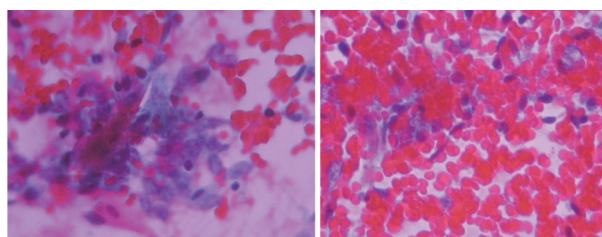


شکل- ۱: عدم توافق کامل، مشاهده گر ۱: AGC، مشاهده گر ۲: HSIL، مشاهده گر ۳: ASC-US، مشاهده گر ۴: SCC



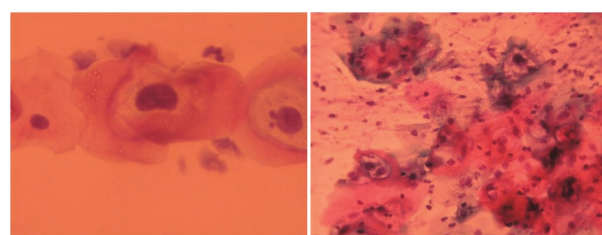


شکل - ۵: توافق کامل AGC



شکل - ۳: توافق نسبی، مشاهده گر ۱ و ۲ و ۴: SCC، مشاهده گر ۳: LSIL

کمتر باشد.<sup>۹</sup> در مطالعه اخیر، همان‌گونه که از جداول پیداست میزان تکرارپذیری در بین مشاهده‌گران در کل اندک بوده  $K=0/26$  و بیشترین میزان تکرارپذیری بین همه مشاهده‌گران مربوط به دسته تشخیصی مهاجمی ( $K=0/53$ ,  $P<0/0001$ ) و کم‌ترین میزان تکرارپذیری بین همه مشاهده‌گران مربوط به دسته تشخیصی HSIL ( $K=0/19$ ,  $P<0/0001$ ) بوده است. بیشترین میزان توافق در بین مشاهده‌گرها بین مشاهده‌گر ۱ (مشاهده‌گر مرجع) و مشاهده‌گر ۲ وجود داشت (مربوط به دسته تشخیصی LSIL) و کم‌ترین میزان توافق کشف شده مربوط به مشاهده‌گر ۳ و ۴ در دسته تشخیصی SCC بود. میانگین کاپا بین مشاهده‌گران ۱ و ۲ و بین مشاهده‌گران ۱ و ۴ برابر بود ( $K=0/46$ ). میزان توافق در گروه مهاجم‌ها کمی بهتر از گروه مهاجم‌ها بود ( $K=0/53$  در مقابل  $K=0/32$ ). توافق در گروه اسکواموس و غددی تفاوت چندانی باهم نداشت و در هر دو گروه اندک بود ( $K=0/27$  در مقابل  $K=0/37$ ). در یکی از تازه‌ترین مطالعات (۲۰۰۹) توسط A.P. Weinfurter Lima و همکارانش<sup>۲</sup> تکرارپذیری اسمیرهای سرویکال توسط شش سیتولوژیست از دو آزمایشگاه در گروه‌های: منفی برای ضایعه اینتراپی تلیال یا بدخیمی (NILM)، ASC، LSIL، HSIL، مهاجمی، SCC، AGC، AIS، آدنوکارسینوم مهاجم و تغییرات مخلوط مورد غربالگری قرار گرفتند. نتایج مؤید این حقیقت بودند که گروه ASC دارای کم‌ترین و گروه (NILM) دارای بیش‌ترین میزان توافق کلی بودند و بعد از آن گروه HSIL از توافق کلی بالایی برخوردار بود. مطالعه حاضر در مقایسه با دیگر مطالعات از نظر پارامترهای مطالعه (حجم نمونه، تعداد پاتولوژیست‌های وارد شده در مطالعه و تعداد دسته‌های تشخیصی بررسی شده) از وضعیت خوبی برخوردار می‌باشد (۶۴۸ پاسخ نهایی برای بررسی هشت دسته تشخیصی). بیشترین میزان تکرارپذیری بین فردی در کل در دسته مهاجم‌ها (SCC+ آدنوکارسینوم) و سپس در دسته LSIL بود، که با



شکل - ۴: توافق کامل LSIL

نمونه‌ها در شکل ۳ نشان داده شده است. در ۱۷ مورد ( $0/5/1$ ) توافق کامل در بین همه پاتولوژیست‌ها وجود داشت و همه مشاهده‌گرها یک تشخیص واحد را مطرح کرده بودند که در شکل‌های ۴ و ۵ دو نمونه از آن‌ها نشان داده شده است.

## بحث

تکرارپذیری بین فردی نقش مهمی در تعیین خطاهای تشخیصی و در مباحث پزشکی قانونی دارد. در نمونه‌های سیتولوژی دهانه رحم نیز تکرارپذیری نقش مهمی را در مدیریت هر یک از ضایعات اینتراپی تیل دهانه رحم توسط پزشک معالج دارد و نقش آن را برای برنامه پیشگیرانه پاپ‌اسمیر می‌توان امری حیاتی دانست.<sup>۷،۸</sup> هر چند مطالعات چندانی با به‌کارگیری حجم نمونه قابل توجه برای بررسی تکرارپذیری در سیتولوژی دهانه رحم صورت نگرفته است، لیکن نتایج همین مطالعات محدود حکایت از تکرارپذیری کلی پایین دارند، البته بسیاری از این مطالعات در سال‌های قبل از ۱۹۹۱ و ۲۰۰۱ صورت گرفته است (یعنی قبل از این‌که سیستم بتسدا دست‌خوش تغییراتی برای بهبودی تکرارپذیری شود).<sup>۳</sup> در مطالعات جدید نتایج توافقات کلی و توافق در گروه‌های مختلف تشخیصی مختلف بوده است، اما همان‌طور که در مطالعه BIRST آمده است، انتظار می‌رود که میزان تکرارپذیری در گروه‌های چالشی‌تر یعنی (ASC-US، ASC-H)

منفی گزارش کردن نمونه‌ها که در بعضی موارد تا حد ۲۳٪ نیز رسید، امری قابل توجه است، به خصوص این‌که یک مورد از نمونه‌های منفی گزارش شده توسط مشاهده‌گر مرجع در دسته SCC قرار داده شده بود. علاوه بر این‌ها، نتایج نشان می‌دهد که توافق بر سر غددی یا اسکواموس بودن ضایعه نیز اندک بوده است (K=۰/۳۷) که اهمیت آن مربوط به مدیریت متفاوت این دو گروه بر مبنای راهنمای American Society of Cervical Colposcopy and Pathology (ASCCP) می‌باشد. در کل اندک بودن میزان توافق بین فردی با برخی مطالعات دیگر از این دست هم‌خوانی داشت هر چند که در مجموع مطالعات چندانی در زمینه تکرارپذیری پاپ‌اسمیر صورت نگرفته است. پیشنهاد می‌شود که در تحقیقات آتی، مطالعات گسترده‌تر با میزان حجم نمونه بالاتر و مشاهده‌گران بیشتر برای رسیدن به داده‌هایی قابل قبول‌تر صورت گیرد تا استناد به جمیع نتایج بتواند راه‌گشایی برای مداخلاتی در جهت کاهش میزان اختلاف نظر در مشاهده‌گران باشد. استفاده از معیارهای روشن‌تر برای تفسیر، تجدیدنظر در سیستم طبقه‌بندی برای استاندارد سازی معیارهای مورفولوژیک و یا به کارگیری ابزارهای آموزشی، مداخلاتی هستند که می‌تواند در بهبود تکرارپذیری تست پاپ‌اسمیر مؤثر باشند.

از سوی دیگر آگاه ساختن صادقانه غیر پاتولوژیست‌ها درباره انتظارات بیش از حد از تکنیک فوق نیز یک قدم مثبت و مهم است. در نهایت تنها می‌توان گفت که برای کمک به پیشبرد اهداف برنامه غربالگری پاپ‌اسمیر چیزی که بسیار بدیهی است ضرورت وجود یک تفسیر تکرارپذیر است و باید برای دسترسی به آن تلاش کرد.

مطالعه BIRST<sup>۹</sup> که در آن LSIL از توافق بیشتری برخوردار بوده است و نیز با نتایج مطالعه Renshaw<sup>۶</sup> که در آن LSIL از توافق بیشتری برخوردار بوده است هم‌خوانی دارد. همان‌گونه که انتظار می‌رود و در مطالعاتی مانند BIRST که توسط Sherman انجام شده<sup>۹</sup> و Weinfurter<sup>۸</sup> نیز ثابت شده است دسته ASC ها که دارای ابهام بیشتری هستند معمولاً از میزان تکرارپذیری کمتری برخوردارند. در مطالعه حاضر نیز دسته تشخیصی ASC-US از تکرارپذیری اندک و ASC-H از تکرارپذیری ضعیف برخوردار بودند. اما برخلاف انتظار، کم‌ترین میزان تکرارپذیری در این مطالعه مربوط به دسته HSIL بود، که می‌تواند بیانگر یک زنگ خطر و ضرورت پرداختن به بررسی بیشتر در این دسته تشخیصی در راستای بهبود تکرارپذیری در آن باشد. البته در مطالعه ALTS نیز میزان توافق در HSIL تنها ۴۷٪ بوده است. مقایسه نتایج مطالعه حاضر با سایر مطالعات در جدول ۵ آمده است. خوشبختانه میزان توافق در دسته مهاجم‌ها (SCC+ آدنوکارسینوم) که از نظر پی‌گیری و درمان بیماران اهمیت دارند از وضعیت مطلوب‌تری نسبت به دیگر دسته‌های تشخیصی برخوردار بود (توافق نسبتاً خوب K=۰/۵۳)، هرچند که میزان توافق در دسته SCC به تنهایی اندک بود. با توجه به این‌که توافق در زمان جمع‌بندی SCC با آدنوکارسینوم افزایش یافت، می‌توان این نتیجه را گرفت که خوشبختانه مشاهده‌گران از نظر مهاجم بودن تشخیص بر روی تعداد بیشتری از اسلایدها توافق داشته‌اند و کمتر بودن توافق در SCC تا حدی به این دلیل بوده است که مشاهده‌گران تعدادی از لام‌هایی با تفسیر اولیه SCC را در دسته تشخیصی آدنوکارسینوم قرار داده بودند. هرچند تکرارپذیری در مرز بین مثبت و منفی امکان‌پذیر نشد ولی

## References

1. Bibbo M, Wilbur D. Comprehensive Cytology. 3<sup>rd</sup> ed. Philadelphia: Saunders; 2008. Part II, Section A. p. 77-87, 184-205.
2. Weinfurter L, Plewka J, Basso J, Torres J, Philipi L, Cicchetti D, et al. Interobserver variability of cervical smears by rescreening in small laboratory of Parana and Santa Catarina, Brazil. *Revista Eletronica de Farmacia* 2009;6(3):36-47.
3. Koss LG. Squamous carcinoma of uterine cervix and its precursors. In: Koss LG, Melamed M. Koss' Diagnostic Cytology and It's Histopathologic Bases. 5<sup>th</sup> ed. Philadelphia, PA: Lippincott, Williams and Wilkins; 2005. p. 306-07, 320-70.
4. Evans DM, Shelley G, Cleary B, Baldwin Y. Observer variation and quality control of cytodiagnosis. *J Clin Pathol* 1974;27(12):945-50.
5. O'Sullivan JP, Ismail SM, Barnes WS, Deery AR, Gradwell E, Harvey JA, et al. Interobserver variation in the diagnosis and grading of dyskaryosis in cervical smears: specialist cytopathologists compared with non-specialists. *J Clin Pathol* 1994;47(6):515-8.
6. Renshaw AA, Davey DD, Birdsong GG, Walsh M, Styer PE, Mody DR, et al. Precision in gynecologic cytologic interpretation: a study from the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. *Arch Pathol Lab Med* 2003;127(11):1413-20.
7. Gupta DK, Komaromy-Hiller G, Raab SS, Nath ME. Interobserver and intraobserver variability in the cytologic diagnosis of normal and abnormal metaplastic squamous cells in pap smears. *Acta Cytol* 2001;45(5):697-703.

8. Stoler MH, Schiffman M; Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA* 2001;285(11):1500-5.
9. Sherman ME, Dasgupta A, Schiffman M, Nayar R, Solomon D. The Bethesda Interobserver Reproducibility Study (BIRST): a web-based assessment of the Bethesda 2001 System for classifying cervical cytology. *Cancer* 2007;111(1):15-25.
10. Gatscha RM, Abadi M, Babore S, Chhieng D, Miller MJ, Saigo PE. Smears diagnosed as ASCUS: interobserver variation and follow-up. *Diagn Cytopathol* 2001;25(2):138-40.
11. Simsir A, Hwang S, Cangiarella J, Elgert P, Levine P, Sheffield MV, et al. Glandular cell atypia on Papanicolaou smears: interobserver variability in the diagnosis and prediction of cell of origin. *Cancer* 2003;99(6):323-30.
12. Cocchi V, Carretti D, Fanti S, Baldazzi P, Casotti MT, Piazzi R, et al. Intralaboratory quality assurance in cervical/vaginal cytology: evaluation of intercytologist diagnostic reproducibility. *Diagn Cytopathol* 1997;16(1):87-92.
13. Renshaw AA, Lee KR, Granter SR. Use of statistical analysis of cytologic interpretation to determine the causes of interobserver disagreement and in quality improvement. *Cancer* 1997;81(4):212-9.
14. Smith AE, Sherman ME, Scott DR, Tabbara SO, Dworkin L, Olson J, et al. Review of the Bethesda System atlas does not improve reproducibility or accuracy in the classification of atypical squamous cells of undetermined significance smears. *Cancer* 2000;90(4):201-6.
15. Woodhouse SL, Stastny JF, Styer PE, Kennedy M, Praestgaard AH, Davey DD. Interobserver variability in subclassification of squamous intraepithelial lesions: Results of the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. *Arch Pathol Lab Med* 1999;123(11):1079-84.
16. Dawson B, Trapp RG. Basic and Clinical Biostatistics. 4<sup>th</sup> ed. New York: Lange Medical Books, McGraw-Hill; 2004.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159-74.

## Determining the interobserver reproducibility of Pap smears in the diagnosis of epithelial cell abnormalities

Received: March 06, 2011 Accepted: April 18, 2011

### Abstract

Narges Izadi-Mood MD.<sup>1\*</sup>  
Soheila Sarmadi MD.<sup>1</sup>  
Farzaneh Heydari-Farzan MD.<sup>2</sup>  
Hayeda Haeri MD.<sup>3</sup>  
Zahra Forouhesh-Tehrani MD.<sup>2</sup>

1- Department of Pathology,  
Women's Hospital, Tehran  
University of Medical Sciences,  
Tehran, Iran.

2- Department of Pathology,  
Shariati Hospital, Tehran  
University of Medical Sciences,  
Tehran, Iran.

3- Department of Pathology, Imam  
Khomeini Hospital, Tehran  
University of Medical Sciences,  
Tehran, Iran.

**Background:** Cancer of uterine cervix is the second cause of death in women in the world and the most common cause in developing countries. Because the majority of women with invasive cervical cancer of the uterine have not previously undergone screening, many clinicians assume that Pap smear has a high degree of accuracy; but problems such as false positive and false negative interpretations, as well as interobserver variability have questioned its validity.

**Methods:** We retrieved 162 positive cervical smears that had been originally interpreted as ASC-US, ASC-H, LSIL, HSIL, SCC, AGC and adenocarcinoma from the cytology archives of Women's Hospital in Tehran, Iran. The slides were rescreened by an experienced pathologist and reclassified in the mentioned categories. All the 162 slides were reviewed by three more pathologists in a blind study using interpretative criteria utilized in their daily routine to evaluate interobserver reproducibility. To increase the level of interobserver agreement, the diagnostic categories were reduced to squamous *Vs.* glandular abnormalities and invasive (SCC and adenocarcinoma) *Vs.* non-invasive abnormalities.

**Results:** The results obtained in this study indicated slight interobserver agreement ( $k=0.26$ ). The most reproducible category was the invasive category (SCC in addition to adenocarcinoma) and the least agreement was seen for HSIL ( $k=0.19$ ).

**Conclusion:** This study showed that reproducibility of cytological interpretation of conventional Pap smears varies among interpretative categories and the overall interobserver agreement is slight. Since convening on the reduction of interobserver discrepancy in Pap smear interpretations necessitates more reliable information of interpretative variability, larger studies need to be undertaken.

**Keywords:** Abnormality, glandular, Pap smear, reproducibility, squamous cell.

\*Corresponding author: Department of Pathology, Women's Hospital Nejatollahi St, Karim Khan-e- Zand Ave, Tehran, Iran.  
Tel: +98-21-88906767  
email: nizadimood@yahoo.com