

بررسی تفاوت بین و درون مشاهده‌گر در تعیین ضایعات التهابی بیماری ریفلکس بر اساس طبقه‌بندی لوس آنجلس

دکتر سیاوش ناصری مقدم^۱، دکتر رضا ملک‌زاده^۲، دکتر هادی راز جویان^۳، دکتر سیدمهدی نورایی^۴، دکتر سیده‌میثم علی‌محمدی^۳، دکتر منصوره ممرآبادی^۲، دکتر همایون واحدی^۴، دکتر اکرم پورشمس^۴، دکتر مهدی محمدنژاد^۵، دکتر فرهاد زمانی^۵، دکتر فرهاد صدر^۶، دکتر صدیف درویش‌مقدم^۷، دکتر پانته آفرسی^۸

^۱ دانشیار، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی تهران
^۲ استاد، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی تهران
^۳ پژوهشگر، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی تهران
^۴ استادیار، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی تهران
^۵ استادیار، مرکز تحقیقات بیماریهای گوارش و کبد، دانشگاه علوم پزشکی ایران
^۶ استادیار، بیمارستان میلاد تهران
^۷ استادیار، دانشگاه علوم پزشکی کرمان
^۸ فوق تخصص بیماریهای گوارش و کبد، بیمارستان تأمین اجتماعی تهران

چکیده

زمینه و هدف

در حال حاضر از طبقه‌بندی لوس آنجلس برای گزارش یکسان مشاهدات آندوسکوپی بیماری ریفلکس استفاده می‌شود. هدف ما از انجام این مطالعه بررسی تغییرپذیری بین و درون مشاهده‌گر طبقه‌بندی لوس آنجلس است.

روش بررسی

۲۵۴ تصویر انتهای مری که توسط یک آندوسکوپیست متبحر گرفته شده بود به‌طور تصادفی انتخاب و توسط ۹ متخصص دستگاه گوارش (۶ متخصص و ۳ دستیار فوق تخصصی) خوانده شد. از این افراد خواسته شد تا مشاهدات خود را بر پایه طبقه‌بندی لوس آنجلس گزارش کنند. بعد از ۲ هفته تصاویر مجدداً کدبندی شده و از همکاران خواسته شد تا دوباره تصاویر را مشاهده کنند. از آنالیز آماری کاپا برای تغییرپذیری بین و درون مشاهده‌گر استفاده شد. مقادیر p کمتر از ۰/۰۵ از نظر آماری معنی دار در نظر گرفته شد.

یافته‌ها

متوسط کاپا برای توافق درون مشاهده‌گر متخصصین ۰/۵۴ و برای دستیاران فوق تخصصی ۰/۵۵ بود. متوسط کاپای درون مشاهده‌گر خانمها بیشتر از آقایان بود (۰/۷۰ در مقابل ۰/۵۰، $p = ۰/۰۵$). متوسط کاپای بین مشاهده‌گران برای متخصصین و دستیاران فوق تخصصی به ترتیب ۰/۲۰ و ۰/۳۱ گزارش شد. متوسط کاپای وزنی بین مشاهده‌گر به ترتیب برای خانمها و آقایان ۰/۰۷ و ۰/۲۵ بود ($p = ۰/۰۰۷$). آنالیز داده‌ها برای شناسایی منبع خطا نشان داد که حداقل تکراری خواندن برای محاسبه ضریب توافق هر دو گروه بین و درون مشاهده‌گر ریفلکس گرید A بود.

نتیجه‌گیری

اگرچه طبقه‌بندی لوس آنجلس پیشرفت مهمی در گزارش بیماری ریفلکس معدی-مروی است، بر پایه داده‌های ما تکرارپذیری این طبقه‌بندی ضعیف تا متوسط است. هیچ تفاوتی بین توافق متخصصین و دستیاران فوق تخصصی مشاهده نشد. کاپای درون مشاهده‌گر خانمها بهتر گزارش شد اما میزان توافق بین آنها کمتر بود. علی‌رغم نواقص آماره کاپا و محدودیتهای احتمالی کیفیت تصاویر ثابت، بازنگری سیستم لوس آنجلس پیشنهاد می‌شود.

کلید واژه: توافق درون مشاهده‌گر، توافق بین مشاهده‌گر، آندوسکوپی، ریفلکس معدی مروی، طبقه‌بندی لوس آنجلس

گوارش / دوره ۱۱، شماره ۱، بهار ۱۳۸۵، ۲۷-۳۲

تاریخ پذیرش: ۸۵/۲/۱۴

تاریخ اصلاح نهایی: ۸۵/۲/۱۲

تاریخ دریافت: ۸۴/۱۲/۴

زمینه و هدف

بیماری ریفلکس معدی-مروی * (GERD) مشکل شایع و مزمنی است (۱)، که میزان بروز آن در طی دو دهه اخیر به‌طور چشمگیری در غرب

نویسنده مسئول: تهران، خیابان کارگر شمالی، بیمارستان شریعتی،

کدپستی ۱۴۱۱۴

تلفن: ۸۸۰۱۲۹۹۲ نمابر: ۸۸۰۱۲۹۹۲

* Gastroesophageal Reflux Disease

E-mail: sianm@ams.ac.ir

طبقه بندی LA به خیلی از اهداف خود رسیده است، اما هر ابزار یا سیستمی باید قبل از مورد قبول واقع شدن تکرار پذیر باشد که این حقیقت در مورد طبقه بندی LA جهت یافته های آندوسکوپی یک بیماری ریفلاکس نیز صادق می باشد. مطالعات در این زمینه اندک هستند.

روش بررسی

در این مطالعه روش بررسی یافته های آندوسکوپی یک قسمت انتهایی مری (طبقه بندی لوس آنجلس) از جهت تغییر پذیری درون و بین مشاهده گر ارزیابی شده که در آن از تصاویر آندوسکوپی آرشیو یک آندوسکوپیست متبحر استفاده شد. ۲۵۴ قطعه عکس از قسمت انتهایی مری با کمک نرم افزار کامپیوتری به طور تصادفی انتخاب شدند. تصاویر توسط چند نفر از پژوهشگران که در گزارش آن تصاویر دخیل نبودند کدبندی شدند. مجموعه تصاویر کدبندی شده بر روی دیسک فشرده ضبط شد. ۹ آندوسکوپیست (۲ نفر متبحر در زمینه GERD، ۴ فوق تخصص گوارش، ۳ دستیار فوق تخصصی گوارش که معمولاً در کارشان از طبقه بندی LA استفاده می کردند دعوت به همکاری شدند. سپس به هر کدام از آنان یک نسخه از تصاویر دیجیتال داده شد و از آنان خواسته شد که تصاویر را بر اساس طبقه بندی LA گزارش نمایند. در هر فرم گزارش علاوه بر وجود ستونهای درجه بندی ازوفاژیت و کد عکس، دو ستون اضافه تحت عنوان موارد طبیعی و ملاحظات نیز موجود بود. بعد از ۲ هفته ترتیب تصاویر عوض شد و کدهای جدید به آنها داده شد و CD جدید در اختیار همان آندوسکوپیست ها قرار گرفت تا تصاویر را گزارش نمایند. از آماره کاپا و کاپای وزنی برای مقایسه تغییر پذیری بین درون مشاهده گر این گزارشها استفاده شد. کاپا مانع از تأثیرگذاری شانس بر توافق به دست آمده در داده های مامی شود و پزشک را از میزان توافق، بدون در نظر گرفتن شانس آگاه می سازد. در صورتی که مشاهده گرها در تک تک موارد توافق داشته باشند حداکثر میزان توافق ۱۰۰ درصد خواهد بود. میزان کاپا از ۱- (عدم توافق کامل) تا ۱+ (توافق کامل) متغیر است و عدد ۰ به شانس بودن این توافق دلالت دارد. کاپای بالای ۰/۸ توافق عالی و زیر ۰/۲ خیلی ضعیف در نظر گرفته شد. مقادیر بین ۰/۶۰-۰/۷۹ خوب، ۰/۴۰-۰/۵۹ متوسط و ۰/۲۰-۰/۳۹ ضعیف می باشند. علاوه بر بررسی هر درجه ازوفاژیت به طور جداگانه در کل گروه، آنالیز زیرگروهی نظیر مقایسه بین دستیار و سایر متخصصین و زیرگروه زنان و مردان نیز انجام شد. بعد از نگاه به اولین سری آنالیز داده ها، آماره کاپا برای گروههای ترکیبی نیز محاسبه شد. برای مثال جهت پیدا کردن منشأ اختلاف، مجموعه نرمال و درجه A به عنوان یک گروه، درجه A و درجه B تحت عنوان یک گروه و ترکیب درجه B و درجه C گروهی دیگر را تشکیل دادند. علاوه بر این، جهت تعیین تأثیر هر

* Los Angeles Classification

افزایش یافته است. (۲)، گزارشهای اخیر از کشورهای در حال توسعه نیز نشان دهنده همین روند می باشد. بیماری ریفلاکس علاوه بر تحمیل بار مالی زیاد بر نظام مراقبت سلامت، روی جنبه های مختلف سلامت زندگی بیماران نظیر کیفیت زندگی آنان نیز تأثیرگذار است. (۴، ۵، ۶، ۷)، به همین دلیل این بیماری در طی دو دهه اخیر بسیار مورد توجه قرار گرفته است. در خیلی از موارد معمولاً از آندوسکوپی برای تشخیص بیماری، دیدن عوارض آن (مری بارت، تنگی، زخم، آدنوکارسینوم)، و رد بیماریهای خوش خیم یا بدخیم همراه استفاده می شود. (۷)، در موارد ازوفاژیت شدید یا عمیق تشخیص آندوسکوپی یک GERD آسان است اما از آنجایی که اکثر بیماران ازوفاژیت شدید ندارند و هر فردی تغییرات خفیف را نمی بیند و گزارش نمی کند، بنابراین تشخیص تغییرات خفیف تر مری در این بیماری مهم است. وجود تفاوت های زیاد بین مشاهده گرها موجب بی ثمر ماندن تلاشهای انجام شده برای یکی کردن مشاهدات توسط سیستمهای طبقه بندی پیشرفته مختلف شده است. تاکنون بیش از ۸۰ مورد سیستم طبقه بندی مختلف برای ازوفاژیت ناشی از ریفلاکس ابداع شده است. از این میان تنها تعداد اندکی از آنان مقبولیت عمومی دارند. در حال حاضر اکثر مطالعات مربوط به اثربخشی داروها در تأثیر درمان GERD از یکی از سیستمهای طبقه بندی مقبولتر برای گزارش التیام ازوفاژیت استفاده می کنند. یکی از رایجترین سیستمهای طبقه بندی آندوسکوپی یک GERD، طبقه بندی لوس آنجلس* (LA) است که از سال ۱۹۹۴ به کار گرفته شده است. (۸)، در طبقه بندی LA تلاش شده است تا از گزارشهای فردی اجتناب و مبتنی بر معیارهای عینی برای تشخیص آندوسکوپی یک GERD شود. در طبقه بندی LA دیدن شکست واضح مخاطی (erosion) برای تشخیص آندوسکوپی یک GERD ضروری است. بر اساس طول شکست و نحوه گسترش آن بین چینهای مری، درجات متفاوتی از ازوفاژیت تعریف شده است (جدول ۱). زخمها و تنگیها در این سیستم جداگانه گزارش می شوند.

جدول ۱: طبقه بندی لوس آنجلس برای ازوفاژیت روزیو ** *

| | |
|--------|--|
| درجه A | شکست مخاطی کمتر یا مساوی ۵ میلی متر، بین چینهای مخاطی گسترده نشده است. |
| درجه B | شکست مخاطی بیش از ۵ میلی متر، بین چینهای مخاطی گسترده نشده است. |
| درجه C | شکست مخاطی بین چینها گسترده شده است اما کمتر از ۷۵٪ محیط مری را درگیر کرده است. |
| درجه D | شکست مخاطی بین چینها گسترده شده است و بیشتر یا مساوی ۷۵٪ محیط مری را درگیر کرده است. |

* خوردگی به صورت ترکهای واضح مخاطی با کنتراست مشخص در مخاط به ظاهر سالم یا بدون آگزودا تعریف شده است.

** زخمها، تنگیها و متاپلازی استوانه ای به صورت مجزا با هر درجه گزارش می شود.

معنی داری نداشت. میانگین کاپای وزنی در هر بار خواندن تصاویر توسط دستیاران فوق تخصصی و سایر متخصصین در جدول ۳ نشان داده شده است.

دستیاران فوق تخصصی گزارش در بار اول خواندن تصاویر مختصری بهتر از استادان بوده اند که این میزان در بار دوم خواندن تصاویر از نظر آماری معنی دار بود ($p = 0/002$). میانگین کلی کاپای وزنی بین مشاهده گر (SD) در دو گروه متخصصین و دستیاران فوق تخصصی گزارش $0/20 (0/13)$ و $0/31 (0/05)$ گزارش شد. میانگین کاپای وزنی بین مشاهده گر در دو گروه زنان و مردان به ترتیب $0/25 \pm 0/09$ و $0/07 \pm 0/07$ ($p = 0/007$) بود. ترکیب GERD-A و نرمال در یک گروه، GERD-B و GERD-A در یک گروه و GERD-C و GERD-B در گروهی دیگر در شناختن منبع خطا کمکی نکرد. هر چند که ترکیب GERD-A و نرمال باعث بهبود مختصر آماره کاپا شد.

توافق درون مشاهده گر (Intra observer agreement)

متوسط کاپای وزنی در تغییرپذیری درون مشاهده گر $0/54$ بود (محدوده $0/40-0/72$ ، $SD = 0/11$). بین دستیاران فوق تخصصی گزارش و متخصصین گزارش تفاوت آماری معنی داری دیده نشد (به ترتیب کاپا: $0/15 \pm 0/50$ در مقابل $0/54 \pm 0/10$).

متوسط کاپای وزنی در دو گروه مردان و زنان در جدول ۴ نشان داده شده است. همان طور که می بینید ثبات خواندن عکسها در هر دو بار در خانمها بیش از آقایان بوده است ($p < 0/0001$). مجدداً ترکیب GERD-A و نرمال، GERD-B و GERD-A، GERD-C و GERD-B در شناختن منبع اختلاف کمکی نکرد، اما کاپای وزنی در گروه GERD-A و نرمال و GERD A,B مختصری بهتر بوده است.

درجه روی کارایی و توافق کلی، درجات مختلف را به ترتیب با هم ادغام کرده و با یک گروه مشخص مقایسه نمودیم.

برای کاپای وزنی از نسخه نهم بسته آنالیز آماری (LP، 2005، شرکت STATA) استفاده شد. از آزمون ویلکاکسون برای مقایسه مقادیر کاپا در هر دو بار خواندن تصاویر استفاده گردید. مقادیر کمتر از $0/05$ از نظر آماری معنی دار در نظر گرفته شد. جدول ۲ وزن درجات مختلف را نشان می دهد که برای محاسبه کاپای وزنی به کار رفته است.

جدول ۲: وزنها تعیین شده برای محاسبه کاپای وزنی

| طبیعی | درجه A | درجه B | درجه C | درجه D |
|--------|--------|--------|--------|--------|
| طبیعی | ۱ | | | |
| درجه A | ۰/۹ | ۱ | | |
| درجه B | ۰/۸ | ۰/۹ | ۱ | |
| درجه C | ۰/۷ | ۰/۸ | ۰/۹ | ۱ |
| درجه D | ۰/۶ | ۰/۷ | ۰/۸ | ۰/۹ |

یافته ها

در این مطالعه ۹ متخصص گزارش شرکت داشتند که ۳ نفر دستیار فوق تخصصی سال دوم گزارش بودند و همه آنها هر دو مرحله مطالعه را به پایان رساندند.

توافق بین مشاهده گر (Inter observer agreement)

میانگین کاپای وزنی برای تغییرپذیری بین مشاهده گر در بار اول خواندن تصاویر $0/22$ (محدوده $0/02-0/49$ ، $SD = 0/11$) و برای بار دوم $0/23$ (محدوده $0/01-0/42$ ، $SD = 0/12$) گزارش شد که اختلاف

جدول ۳: کاپای وزنی بین مشاهده گر (بار اول خواندن با قلم پررنگ و بار دوم با قلم ایتالیک مشخص شده است). دستیاران گزارش با علامت ستاره نشاندار شده اند. موارد F و I خانم می باشند.

| I* | H* | G* | F | E | D | C | B | A |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------|
| <i>0/327</i> | <i>0/287</i> | <i>0/290</i> | <i>0/106</i> | <i>0/282</i> | <i>0/237</i> | <i>0/451</i> | <i>0/238</i> | A |
| <i>0/254</i> | <i>0/251</i> | <i>0/245</i> | <i>0/024</i> | <i>0/259</i> | <i>0/047</i> | <i>0/190</i> | | B |
| <i>0/292</i> | <i>0/397</i> | <i>0/245</i> | <i>0/153</i> | <i>0/319</i> | <i>0/318</i> | | <i>0/154</i> | C |
| <i>0/121</i> | <i>0/203</i> | <i>0/140</i> | <i>0/497</i> | <i>0/059</i> | | <i>0/428</i> | <i>0/106</i> | D |
| <i>0/299</i> | <i>0/326</i> | <i>0/326</i> | <i>0/040</i> | | <i>0/183</i> | <i>0/173</i> | <i>0/289</i> | E |
| <i>0/069</i> | <i>0/092</i> | <i>0/066</i> | | <i>0/028</i> | <i>0/184</i> | <i>0/151</i> | <i>0/017</i> | F |
| <i>0/251</i> | <i>0/274</i> | | <i>0/028</i> | <i>0/243</i> | <i>0/133</i> | <i>0/149</i> | <i>0/362</i> | G |
| <i>0/325</i> | | <i>0/315</i> | <i>0/027</i> | <i>0/422</i> | <i>0/213</i> | <i>0/197</i> | <i>0/313</i> | H |
| | <i>0/345</i> | <i>0/394</i> | <i>0/072</i> | <i>0/356</i> | <i>0/302</i> | <i>0/338</i> | <i>0/254</i> | I |

جدول ۴: کاپای وزنی تغییر پذیری درون مشاهده‌گر. دستیاران گوارش با علامت ستاره نشاندار شده‌اند. موارد F و I خانم می‌باشند.

| I* | H* | G* | F | E | D | C | B | A | |
|------|------|------|------|------|------|------|------|------|------------|
| ۰/۷۲ | ۰/۵۰ | ۰/۴۳ | ۰/۶۹ | ۰/۴۱ | ۰/۶۳ | ۰/۵۴ | ۰/۵۰ | ۰/۵۰ | به‌طور کلی |
| ۰/۸ | ۰/۴۲ | ۰/۳۷ | ۱ | ۰/۳۴ | ۰/۲۷ | ۰/۴۳ | ۰/۵۲ | ۰/۵۶ | طبیعی و A |
| ۰/۷ | ۰/۵۵ | ۰/۴۵ | ۰/۶۱ | ۰/۴۲ | ۰/۶۷ | ۰/۴۴ | ۰/۴۵ | ۰/۴۶ | A & B |
| ۰/۷۱ | ۰/۵۱ | ۰/۴۴ | ۰/۶۹ | ۰/۴۳ | ۰/۶۳ | ۰/۵۷ | ۰/۵۳ | ۰/۵۰ | B & C |
| ۰/۶۵ | ۰/۵۶ | ۰/۴۷ | ۰/۶۱ | ۰/۴۷ | ۰/۶۹ | ۰/۷۹ | ۰/۴۹ | ۰/۴۷ | طبیعی |
| ۰/۵۹ | ۰/۴۲ | ۰/۲۵ | ۰/۴۹ | ۰/۲۹ | ۰/۶۵ | ۰/۵۱ | ۰/۲۶ | ۰/۳۸ | A |
| ۰/۶۸ | ۰/۲۶ | ۰/۳۰ | ۱ | ۰/۲۳ | ۰/۴۰ | ۰/۳۶ | ۰/۶۲ | ۰/۶۱ | B |
| ۰/۷ | ۰/۴۸ | ۰/۲۱ | --- | ۰/۰۹ | --- | ۰/۱۴ | ۰/۰۲ | ۰ | C |

بحث

ما در این مطالعه جهت بررسی تکرار پذیری بین و درون مشاهده‌گر طبقه‌بندی LA از ۹ متخصص گوارش (۳ دستیار فوق تخصصی گوارش سال دوم و ۶ متخصص) برای خواندن تصاویر الکترونیکی انتهای مری در دو مرحله درخواست همکاری کردیم. از آماره کاپا برای آنالیز داده‌ها استفاده شد. آماره کاپا و کاپای وزنی می‌توانند میزان توافق با حذف شانس را پیش‌بینی کنند. (۹، ۱۰)، بر پایه اطلاعات ما تکرار پذیری درون مشاهده‌گر طبقه‌بندی LA متوسط است، در حالی که تکرار پذیری بین مشاهده‌گر آن نسبتاً ضعیف می‌باشد. این موضوع خصوصاً در مورد درجات پایین ازوفازیت صادق است.

در مطالعات ژاپنیها دستیاران فوق تخصصی گوارش بدتر از متخصصین کار کردند و زمانی که گزارشها به پایان رسید نشان داده شد که بهتر است طبقه‌بندی LA را متخصصین گوارشی که متبحرتر می‌باشند استفاده کنند. مطالعه ما تفاوتی بین دستیاران و متخصصین نشان نداد که می‌تواند ناشی از تفاوت سیستمهای آموزشی ایران و ژاپن، دفعات متفاوت تماس با بیماری به لحاظ موقعیت جغرافیایی دو کشور یا گزارش متفاوت سیستم لوس آنجلس در هر دو کشور باشد. در مطالعه ما دو متخصص گوارش خانم همکاری داشتند که هر دو نفر آنها کار خواندن تصاویر را در هر ۲ مرحله به نواحی انجام دادند (۰/۷۲ و $K=0/70$). اما زمانی که آنها را از نظر کاپای بین مشاهده‌گر مقایسه کردیم با عددی نسبتاً پایین مواجه شدیم ($K=0/07$). بنابراین متخصصین گوارش خانم ازوفازیت را بهتر می‌بینند اما روشی که هر کدام می‌بینند با دیگری متفاوت است. در مورد اینکه چرا خانمها ازوفازیت را بهتر گزارش کرده‌اند، فرضیه‌ای که می‌توان مطرح کرد این است که خانمها جزئیات را بهتر می‌بینند و قوانین مشاهده را سخت‌گیرانه‌تر اجرا می‌کنند.

GERD مشکلی شایع و در حال افزایش است. از یافته‌های آندوسکوپییک ازوفازیت جهت تأیید تشخیص و همچنین به‌عنوان نقطه نهایی درمان موفق این بیماری با داروهای جدید استفاده می‌شود. بنابراین ضروری است که ابزاری معتبر و با دقت مناسب برای تشخیص آندوسکوپییک ازوفازیت داشته باشیم. همچنین برای دقیق بودن ابزار لازم است استفاده از آن آسان و در میان مشاهده‌گرهای مختلف نتایج مشابهی داشته باشد. سیستمهای طبقه‌بندی مختلفی برای نایل شدن به این هدف ارائه شده‌اند که هر کدام از آنها مقبولیت مختص خود را دارند. از میان سیستمهای طبقه‌بندی که در دنیا بیش از همه مورد استفاده قرار می‌گیرند می‌توان به سیستمهای ساواری-میلر، MUSE* و لوس آنجلس اشاره کرد. هر کدام از این سیستمهای طبقه‌بندی مزایا و محدودیتهای خاص خود را دارند. یکی از مهمترین مشکلات کار با سیستمهای قدیمی تر این است که در آنها تغییرات مختصر (مثل قرمزی مخاط، شکستگی) به‌عنوان درجه خفیفی از ازوفازیت در نظر گرفته می‌شوند که این تغییرات تعریف بسیار مبهمی دارند.

از طرف دیگر بیشتر بیماران مبتلا به GERD شکست مخاطی ندارند و یا تنها درجات خفیفی از روزیون در انتهای مری دارند، بنابراین مهم است که بتوانیم درجات خفیف تر ازوفازیت را با اطمینان و دقت گزارش کنیم. پیشرفت طبقه‌بندی لوس آنجلس با تکیه بر گزارش شکستهای واضح (ترک مخاطی یا یادون آگزودا) بر این مشکلات فائق آمده است. شکست واضح نشانه‌ای از بیماری ریفلاکس است. از این رو با کنار گذاشتن تغییرات مختصر، این طبقه‌بندی به‌عنوان تکرارپذیرترین سیستم طبقه‌بندی برای گزارش تغییرات آندوسکوپییک GERD در نظر گرفته شد.

* Metaplasia, Ulceration, Stricture, Erosions

باشد؛ در غیر این صورت حتی خطاهای کوچک اندازه‌گیری به دلیل شیوع بالای بیماری باعث ارقام نادرست بزرگی می‌شود.

نتیجه‌گیری

بر پایه این مستندات گزارش درجات پایین ازوفاژیت طبق طبقه‌بندی لوس آنجلس بین مشاهده‌گرهای مختلف تکرارپذیر نمی‌باشد که عوامل مختلفی ممکن است پاسخگوی این امر باشند:

(۱) ما از تخمین چشمی برای اندازه‌گیری طول شکستها استفاده می‌کنیم. در طبقه‌بندی لوس آنجلس از طول کمتر یا بیشتر از ۵ میلی‌متر برای افتراق درجات A و B استفاده می‌شود. تخمین چشمی این تغییرات کوچک بدون استفاده از ابزار اندازه‌گیری می‌تواند باعث خطای طبیعی شود. همین اصل نیز در افتراق بین فقدان شکست از شکستهای کوچک ۲-۱ میلی‌متری (افتراق نرمال از درجه A) دیده می‌شود.

(۲) این اندازه‌گیری در ارگانی انجام می‌شود که متحرک است که به خودی خود اندازه‌گیری را مشکل‌تر می‌سازد.

سؤال مهم دیگری که باید پاسخ داده شود این است که تفاوت‌های بین سیر طبیعی بیماران مبتلا به GERD-A و GERD-B چیست؟ آیا بیماران مبتلا به درجات پایین ازوفاژیت مستعد همان عوارضی هستند که درجات شدیدتر ازوفاژیت استعداد ابتلای بیشتری به آن را دارند (GERD C,D)؟ چنانچه درجات پایین ازوفاژیت از نظر سیر طبیعی فرقی باهم ندارند چرا آنها را از هم جدا می‌کنیم؟

با در نظر گرفتن این نکات و داده‌های مطالعه ما زمان مناسبی است که نگاهی مجدد به طبقه‌بندی لوس آنجلس بیاندازیم و کارایی آن را افزایش دهیم.

* Pandolifino

** Rath

مادر این مطالعه از تصاویر انتهایی مری استفاده کردیم که این امر خود موجب اشکال در کیفیت تصاویر شده و می‌تواند روی گزارشها تأثیرگذار باشد. چنانچه ما از تصاویر با زمان حقیقی (real-time) با مدت زمان کوتاه (۱۵-۱۰ ثانیه) استفاده می‌کردیم ممکن بود نتایج متفاوتی به دست آوریم. به طور مثال میانگینهای بهتری در مورد خواندن عکسها برای مشاهده‌گرها حاصل می‌شد. هرچند که این موضوع روی نتایج ما تأثیرگذار است اما میزان این تأثیر مشخص نمی‌باشد. سایر پژوهشگران نیز از تصاویر ثابت و نتایج آنها استفاده کرده بودند و ضریب توافق بهتر از آنچه ما پیدا کردیم به دست آوردند. در مطالعه‌ای که توسط پاندولی‌فینو* و همکاران انجام شد ۹ آندوسکوپیست تصاویر را مرور کردند و دریافتند که متخصصین متبحر بهتر از دستیاران فوق تخصصی گوارش از عهده این کار برآمده‌اند. هرچند مقدار کاپای آنان برای تغییرپذیری بین و درون مشاهده‌گر بیش از کاپای مطالعه ما بود (۵۵/۰ تا ۳۹/۰)، اما هنوز تا توافق مناسب فاصله داشت. (۱۱)

رath* و همکاران از ویدئوکلپ برای ارزیابی تغییرپذیری بین و درون مشاهده‌گر سیستمهای نمره‌بندی مختلف بیماری ریفلکس از جمله LA استفاده کردند. (۱۲)، این محققین بهترین میزان توافق را گزارش کردند (۶۵/۰ تا ۴۹/۰). (K)

بیماری ریفلکس معدی-مروی در کشورهای غربی در حال افزایش است. این بیماری در کشورهای در حال توسعه نیز بیشتر و بیشتر شناخته می‌شود. (۱۴، ۱۳)، در آخرین بررسیهای آندوسکوپی انجام شده در ایران ۳۷ درصد جمعیت ازوفاژیت داشتند. (۱۳)، در مطالعات متاآنالیز اپیدمیولوژیک اخیر شیوع این بیماری در جمعیت عمومی ۲۰-۵ درصد گزارش شده است. (۳)، این داده‌ها نمایانگر شیوع رو به رشد بیماری ریفلکس معدی-مروی است. لذا هر ابزاری که برای ارزیابی جنبه‌های مختلف این بیماری به کار می‌رود باید تا حد امکان دقت بالایی داشته

References

1. Locke GR 3rd, Talley NJ, Fett SL, Zinsmeister AR, Melton LJ 3rd. Prevalence and clinical spectrum of gastroesophageal reflux: a population based study in Olmsted County, Minnesota. *Gastroenterology* 1997; 112: 1448-56.
2. El-Serag HB, Sonnenberg A. Opposing time trends of peptic ulcer and reflux disease. *Gut* 1998; 43: 327-33.
3. Dent J, El-Serag HB, Wallander M-A, Johansson S. Epidemiology of gastroesophageal reflux disease: A systematic review. *Gut* 2005; 54: 710-17.
4. Nasseri-Moghaddam S, Malekzadeh M, Sotoudeh M, Tavangar M, Azimi K, Sohrabpour AA, et al. Lower esophagus in dyspeptic Iranian patients: a prospective study. *J Gastroenterol Hepatol* 2003; 18: 315-21.
5. Wicklund IK, Glise H. Quality of life in different gastrointestinal conditions. *Eur J Surg Suppl* 1998; 582: 56-61.
6. Revicki DA, Wood M, Maton PN, Sorensen S. The impact of gastroesophageal reflux disease, on health-related quality of life. *Am J Med* 1998; 104: 252-8.
7. Armstrong D. Endoscopic evaluation of gastro-esophageal reflux disease. *Yale J Biol Med* 1999; 72: 93-100.
8. Lundell LR, Dent J, Bennett JR, Blum AL, Armstrong D, Galmische JP, et al. Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of

- the Los Angeles classification. *Gut* 1999; 45: 172-80.
9. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *Am J Epidemiol* 1987; 126: 161-9.
 10. Kusano M, Ino K, Yamada T, Kawamura O, Toki M, Ohwada T, *et al.* Interobserver variation in endoscopic assessment of GERD using the "Los Angeles" classification. *Gastrointest Endosc* 1999; 49: 700-4.
 11. Pandolfino JE, Vakil NB, Kahrilas PJ. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists. *Gastrointest Endosc.* 2002; 56: 639-43.
 12. Rath HC, Timmer A, Kunkle C, Endlicher E, Grossmann J, Hellerbrand C, *et al.* Comparison of interobserver agreement for different scoring systems for reflux esophagitis: Impact of level of experience. *Gastrointest Endosc* 2004; 60: 44-9.
 13. Saidi F, Malekzadeh R, Sotoudeh M, Derakhshan MH, Farahvash MJ, Yazdanbod A, *et al.* Endoscopic esophageal cancer survey in the western part of the Caspian Littoral. *Diseases of the Esophagus* 2002; 15: 214-8.
 14. Malekzadeh R, Nasser-Moghaddam S, Sotoudeh M. Gastroesophageal Reflux Disease: The new epidemic. *Arch Iranian Med* 2003; 6: 127-40.

Nasseri-Moghaddam S

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Malekzadeh R

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Razjouyan H

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Nouraei SM

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Alimohammadi SM

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Mamarabadi M

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Vahedi H

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Pourshams A

Digestive Disease Research Center,
Shariati Hospital, Tehran University
of Medical Sciences

Mohamadnejad M

Gastrointestinal and Liver Disease
Research Center, Iran University of
Medical Sciences

Zamani F

Gastrointestinal and Liver Disease
Research Center, Iran University of
Medical Sciences

Sadr F

Tehran Milad Hospital

Darvish-Moghaddam S

Kerman University of Medical
Sciences

Farsi P

Tehran Social Services Hospital

Corresponding Author:

Siavosh Nasseri-Moghaddam M.D.,
Digestive Disease Research Center,
Shariati Hospital, Kargar Ave.,
P.O.Box: 14114, Tehran, Iran.
Telefax: +98 21 88012992
E-mail: sianm@ams.ac.ir

Assessment of Inter-and Intra-Observer Variability of the LA Classification for GERD

ABSTRACT

Background: Los Angeles classification is widely adopted as a means of unifying endoscopic observation for GERD. We assessed the inter- and intra-observer variability of LA classification.

Materials and Methods: Two-hundred fifty-four still images of the lower esophagus taken by an expert endoscopist (RM) were randomly selected and presented to 9 gastroenterologists (2 females, 6 experts, 3 trainees). They were asked to report the images according to LA classification. After 2 weeks the images were re-ordered and the same people were asked to report them again. Kappa-statistics was calculated for intra- and inter-observer variability.

Results: Mean kappa for intra-observer agreement was 0.54 for experts and 0.55 for trainees ($p = \text{NS}$). Mean intra-observer kappa was greater among females than males (0.70 vs 0.50 respectively, $p=0.05$). Mean inter-observer kappa was 0.20 and 0.31 for experts and trainees respectively ($p = \text{NS}$). Mean inter-observer weighted kappas were 0.25 and 0.07 ($p = 0.007$) for males and females respectively. Analyzing data for source of the discrepancy showed that the least reproducible reading was GERD-A both for intra- and inter-observer agreement calculations.

Conclusions: According to our data, the LA classification, although a major advance in reporting GERD, has a poor to fair reproducibility. There was no difference between experts and trainees in using the LA classification. Females seem to be more consistent in their readings, but have less agreement with others. Despite the inherent short-comings of kappa statistics and the limitation caused by the possible quality of the still images, revision of the LA system seems to be advisable. *Govaresh/* Vol. 11, No. 1, Spring 2006; 27-32

Keywords: Inter observer agreement, Intra observer agreement, Endoscopy, GERD, LA classification