



ارائه شده توسط:

سایت ترجمه فا

مرجع جدیدترین مقالات ترجمه شده

از نشریات معتبر

# چک لیست COSMIN برای ارزیابی کیفیت روش شناختی مطالعات در زمینه ویژگی های اندازه گیری ابزارهای اندازه گیری وضعیت سلامت: یک مطالعه بین المللی دلفی

چکیده:

زمینه و هدف: هدف مطالعه COSMIN (استانداردهای مبتنی بر اجماع برای انتخاب ابزارهای اندازه گیری وضعیت سلامت)، توسعه یک چک لیست مبتنی بر اجماع جهت اندازه گیری کیفیت روش شناختی مطالعات مربوط به ویژگی های اندازه گیری بود. ما چک لیست COSMIN و توافق هیئت در زمینه آیتم های مربوط به چک لیست را ارائه می کنیم.

روش ها: یک مطالعه دلفی چهار مرحله ای با کارشناسان بین المللی (روان شناسان، اپیدمیولوژیست ها، متخصصان آمار و پزشکان) انجام شد. از 91 کارشناس دعوت شده، 57 نفر رضایت خود را برای شرکت در مطالعه اعلام کردند (63 درصد). از اعضای هیئت خواسته شد تا موافقت یا مخالفت خود را در خصوص هر طرح پیشنهادی بر روی یک مقیاس پنج نقطه ای رتبه بندی کنند. اجماع زمانی حاصل شد که حداقل 67 درصد از اعضای هیئت، عبارت "موافقم یا کاملاً موافقم" را انتخاب کردند.

نتایج: با استفاده از ویژگی های اندازه گیری زیر، اجماع حاصل شد: همسانی درونی، پایایی، خطای اندازه گیری، روایی محتوایی (از جمله روایی صوری)، روایی ساختار (از جمله روایی ساختاری، آزمون فرضیات و روایی بین فرهنگی)، روایی معیار، حساسیت و تفسیر پذیری. تفسیر پذیری به عنوان یک ویژگی اندازه گیری در نظر گرفته نشد. هم چنین هیئت در خصوص شیوه ارزیابی این ویژگی ها به توافق رسید.

نتیجه گیری: چک لیست COSMIN حاصله در زمان انتخاب یک ابزار اندازه گیری، داوری مقاله، طراحی یا گزارش یک مطالعه در زمینه ویژگی های اندازه گیری و یا برای اهداف آموزشی می تواند مفید باشد.

لغات کلیدی: روش دلفی، ارزیابی پیامد، روان سنجی، کیفیت زندگی، پرسشنامه

مقدمه

اندازه گیری پیامدهای سلامت در تحقیقات علمی و در عملیات بالینی امری ضروری است. بر اساس امتیازات بدست آمده با ابزارهای اندازه گیری، تصمیمات در مورد کاربرد تست‌های تشخیص و درمان گرفته می‌شوند. ابزارهای اندازه گیری وضعیت سلامت بایستی دارای روایی و پایایی (معتبر و قابل اطمینان) باشند. در غیر این صورت، خطر جدی وجود نتایج غیر دقیق و اریب وجود دارد که می‌تواند منجر به نتایج غلطی شود. سازمان‌هایی نظیر سازمان غذا و داروی امریکا (FDA) و آژانس پزشکی اروپا (EMA) بر ابزارهای اندازه گیری دارای روایی خوب تاکید کرده‌اند (1-2). نیاز به ابزارهای اندازه گیری پیامدهای سلامت قابل اطمینان و معتبر (دارای روایی و پایایی) توسط مارشال به خوبی اثبات شده است به طوری که او در کارازمایی‌های شیزوفرنی نشان داد که وقتی ابزار اندازه گیری منتشر نشده در مقایسه با ابزارهای منتشر شده استفاده می‌شوند، نویسندگان و محققان تمایل دارند گزارش کنند که درمان برتر از کنترل است (3).

قبل از استفاده از ابزار اندازه گیری وضعیت سلامت در عملیات بالینی یا تحقیقات، ویژگی‌های اندازه گیری آن یعنی پایایی، روایی و حساسیت، را بایستی به طور کافی ارزیابی کرد. مطالعاتی که به ارزیابی ویژگی‌های اندازه گیری می‌پردازند، بایستی از کیفیت روش شناختی بالایی برای تضمین نتایج مناسب در خصوص ویژگی‌های اندازه گیری یک ابزار اندازه گیری برخوردار باشند.

به منظور ارزیابی کیفیت روش شناختی یک مطالعه بر روی ویژگی‌های اندازه گیری، استانداردهایی لازم است. اگرچه بسیاری از استانداردها و معیارها پیشنهاد شده‌اند، این‌ها به چک لیست‌های کاربرپسند و قابل کاربرد، تبدیل نشده‌اند (4-5). به علاوه، این استانداردها توجه کافی را به مطالعاتی که از مدل "نظریه پاسخ سؤال" استفاده می‌کنند و یا مبتنی بر اجماع نیستند (6-7) مبذول نمی‌کنند. این چک لیست بایستی در بر گیرنده یک مجموعه کاملی از استانداردها (که اشاره به نیازهای طراحی و روش‌های آماری مطلوب دارد) و معیارهای کفایت برای ایجاد ویژگی‌های اندازه گیری خوب باشد. اجماع گسترده و زیادی برای دست یابی به پذیرش گسترده یک چک لیست لازم است.

تحقیقات در زمینه ویژگی‌های اندازه گیری از اهمیت زیادی برای پیامدهای سلامتی که مستقیماً توسط بیماران گزارش می‌شوند یعنی پیامدهای مرتبط با سلامت گزارش شده توسط بیمار (HR-PRO) برخوردار است. HR-

PRO به معنی اندازه گیری هر بهدی از وضعیت سلامت بیمار می باشد که به طور مستقیم توسط بیمار یعنی بدون تفسیر پاسخ های بیمار توسط پزشک یا هر کس دیگر (2) ارزیابی می شود. شیوه های جمع آوری داده ها برای ابزارهای HR-PRO شامل ابزارهای مصاحبه ای، ابزارهای خودکار و یا ابزارهای کامپیوتری می باشند (2). نمونه هایی از HR-PRO شامل پرسشنامه های ارزیابی علایم، وضعیت عملکردی و کیفیت زندگی مرتبط با سلامت می باشند. این ها ساختارهایی هستند که به طور مستقیم قابل اندازه گیری نمی باشند. به دلیل ماهیت ذهنی این ساختارها، ارزیابی این که آیا ابزارهای اندازه گیری این ساختارها را به شیوه ای مطمئن و معتبر اندازه گیری می کنند یا خیر، بسیار مهم است.

هدف رویکرد COSMIN (استانداردهای مبتنی بر اجماع برای انتخاب ابزارهای اندازه گیری سلامت)، بهبود انتخاب ابزارهای اندازه گیری سلامت می باشد. هدف این مطالعه به عنوان بخشی از این طرح، توسعه یک چک لیست حاوی استانداردهایی برای ارزیابی کیفیت روش شناختی مطالعات بر روی ویژگی های اندازه گیری است. این چک لیست به صورت همکاری بین رشته ای و بین المللی با همه کارشناسان مرتبط توسعه یافت. ما یک مطالعه دلفی را برای پاسخ به دو سؤال تحقیق انجام دادیم.

1- کدام ویژگی های اندازه گیری بایستی در چک لیست گنجانده شوند؟

2- چگونه این ویژگی های اندازه گیری بایستی از حیث طرح مطالعه و تحلیل آماری ارزیابی شوند (یعنی استانداردها)؟

در این مقاله، ما یک چک لیست COSMIN را ارائه کرده و به توصیف توافق هیئت کارشناسان در خصوص آیتم های گنجانده شده در چک لیست می پردازیم.

روش ها

هدف چک لیست COSMIN

چک لیست COSMIN بر ارزیابی کیفیت روش شناختی مطالعات در زمینه ویژگی‌های اندازه گیری HR-PRO متمرکز است. ما بر HR-PRO به دلیل پیچیدگی این ابزارها تاکید می‌کنیم. این ابزارها، ساختارهایی را اندازه گیری می‌کنند که هم چند بعدی می‌باشند و هم مستقیماً قابل اندازه گیری نیستند.

به علاوه، ما بر کاربردهای ارزش گذارانه ابزارهای HR-PRO یعنی کاربردهای طولی ارزیابی اثرات درمان یا تغییرات در سلامت به مرور زمان تاکید داریم. عبارت "ارزش گذارانه" لازم است زیرا ملزومات ویژگی‌های اندازه گیری با توجه به کاربرد ابزار متغیر هستند (8). برای مثال، ابزارهای مورد استفاده برای ارزیابی بایستی پاسخ‌گو<sup>1</sup> باشند، در حالی که ابزارهای مورد استفاده برای تمایز و تفکیک این ویژگی را ندارند.

کمیته اجرایی COSMIN (پیوست 1) مرور منابعی را برای تعیین شیوه ارزیابی عمومی ویژگی‌های اندازه گیری انجام داده است. دو جست و جو در این رابطه انجام شد: 1- مرور منابع سیستماتیک برای شناسایی همه مقالات مروری سیستماتیک موجود در خصوص ویژگی‌های اندازه گیری ابزارهای اندازه گیری وضعیت سلامت (9) انجام شد. با توجه به این مرور منابع، اطلاعاتی استخراج شد که بر اساس آن ویژگی‌های اندازه گیری ارزیابی شده و استانداردهایی برای بررسی ویژگی‌های اندازه گیری مطالعات استفاده شدند. برای هر ویژگی اندازه گیری، استانداردهای مختلفی یافته شد که برخی از آنها متناقض می‌باشند (9). 2- کمیته اجرایی نیز مرور منابع سیستماتیک دیگری را برای شناسایی مقالات روش شناسی و متون حاوی استانداردها و معیارهای ارزیابی ویژگی‌های اندازه گیری ابزارهای اندازه گیری وضعیت سلامت انجام داد. مقالات در صورتی انتخاب شدند که هدف مقاله، ارائه چک لیست یا استانداردهایی برای خواص اندازه گیری باشد. استانداردهای شناسایی شده در منابع فوق به صورت ورودی در مراحل دلفی استفاده شد.

### مطالعه بین المللی دلفی

سپس، یک مطالعه دلفی انجام شد که دارای چهار مرحله کتبی بود. اولین پرسش نامه در مارس 2006 ارسال شد و آخرین پرسش نامه در نوامبر 2007 ارسال شد. ما تصمیم گرفتیم تا حداقل 80 کارشناس بین المللی را برای شرکت در هیئت کارشناسان دلفی برای اطمینان از حضور 30 پاسخگو در آخرین مرحله دعوت کنیم. بر

---

<sup>1</sup> responsive

اساس تجربه‌های قبلی در زمینه مطالعات دلفی (10-11)، ما انتظار داشتیم که 70 درصد افراد دعوت شده در مطالعه شرکت کنند و از این تعداد 65 درصد اولین لیست را کامل کردند. در ابتدا انتظار می‌رفت که 75 درصد در مطالعه باقی بمانند. ما کارشناسانی را از رشته‌های روان‌شناسی، اپیدمیولوژی، آمار و پزشکی بالینی جمع‌آوری کردیم. از افراد دعوت شده، نویسندگان مقالات مروری، مقالات روش‌شناختی و متون کتاب حضور داشتند. کارشناسان بایستی حداقل 5 مقاله را در زمینه روش‌های اندازه‌گیری وضعیت سلامت در PubMed می‌داشتند. ما از افراد در بخش‌های مختلف دنیا دعوت به عمل آوردیم.

در مرحله اول، سوالاتی در رابطه با این که کدام ویژگی‌های اندازه‌گیری در چک لیست استفاده شوند و در مورد اصطلاحات و تعاریف آن‌ها پرسیدیم. برای مثال، ما در خصوص پیوستگی درونی ویژگی اندازه‌گیری این سؤال را پرسیدیم که از نظر شما چه اصطلاحی برای این ویژگی اندازه‌گیری بهترین است؟ و گزینه‌های پاسخ شامل "همسانی درونی، پایایی همسانی درونی، همگنی، همسانی مقیاس درونی، پایایی دو نیم‌سازی، پایایی درونی، پایایی ساختاری، همسانی گویه، پایایی بین گویه ای و یا سایر" بودند و مقداری فضای بیشتر برای ارائه یک اصطلاح دیگر وجود داشت. در خصوص تعاریف، ما پرسیدیم که "کدام تعریف برای همسانی درونی مناسب است؟ و هفت تعریف موجود در منابع و گزینه "سایر" به آن‌ها ارائه شد به طوری که اعضای هیئت کارشناسان بتوانند یک تعریف دیگر ارائه کنند. در مرحله دوم، ما سوالاتی را در مورد استانداردهای مطلوب برای هر ویژگی اندازه‌گیری معرفی کردیم. ما سوالاتی در مورد مسائل طراحی یعنی "آیا شما با ملزومات زیر برای طراحی یک مطالعه ارزیابی‌کننده همسانی درونی ابزارهای HR-PRO در کاربرد ارزش‌گذارانه موافقت می‌کنید؟ پرسیدیم. 1- یک مدیریت بایستی موجود باشد 2- یک کنترل برای تک بعدی بودن به ازای هر مقیاس بایستی انجام شود 3- آمار همسانی درونی بایستی به طور جداگانه برای هر زیرمقیاس محاسبه شود. هیئت می‌تواند هر گویه را بر روی یک مقیاس 5 نقطه‌ای متغیر از کاملاً مخالفم تا کاملاً موافقم پاسخ دهد. سپس، از هیئت خواسته شد تا ارزیابی کند که کدام روش‌های آماری برای ارزیابی ویژگی اندازه‌گیری کافی هستند. یک فهرستی از روش‌های آماری مرتبط برای هر ویژگی اندازه‌گیری ارائه شد. برای مثال، برای همسانی درونی، رایج‌ترین روش‌های مورد استفاده پیشنهاد شدند: آلفای کرونباخ، فرمول-20 کودر-ریچاردسون، همبستگی کل-گویه میانگین، همبستگی بین گویه ای کل، تحلیل دو نیم شدن، خوبی برازش در یک سطح جهانی یعنی شاخص تفکیک، خوبی برازش در سطح محلی،

یعنی تست‌های گویه‌های خاص یا سایر. اعضای هیئت کارشناسان می‌توانند بیش از یک روش را نشان دهند. در سومین مرحله، ما رایج‌ترین روش مورد استفاده را ارائه کردیم که هر دو بر اساس CTT و IRT می‌باشند و این سؤال پرسیده شد که آیا هیئت کارشناسان، این روش را مطلوب‌ترین روش برای ارزیابی ویژگی اندازه‌گیری می‌دانند یا خیر. برای همسانی درونی، الفای کرونباخ و خوبی برازش در سطح عمومی یعنی شاخص تفکیک در نظر گرفته شد. در سومین مرحله، از اعضای هیئت پرسیده شد که آیا روش‌های دیگر (فرمول 20 کور-ریچاردسون، همبستگی کل گویه‌ها، همبستگی بین گویه‌ای میانگین، تحلیل دو نیم شدن، خوبی برازش در سطح محلی مناسب هستند یا خیر؟ اعضای هیئت هم چنین "سایر روش‌ها" را در مرحله 2 نشان دادند. روش‌های نشان داده شده شامل مقادیر ویژه و یا درصد واریانس توجیه شده از تحلیل عاملی، موکن رو یا لایونگر H برای همسانی درونی بودند. در مرحله 3، از هیئت پرسیده شد که آیا آن‌ها این روش‌ها را برای ارزیابی همسانی درونی مناسب می‌دانند یا خیر. در مرحله دلفی نهایی، همه ویژگی‌ها و استانداردهای اندازه‌گیری که مورد توافق هیئت بود توسط کمیته اجرایی به نسخه مقدماتی چک لیست برای ارزیابی کیفیت روش شناختی مطالعات بر روی ویژگی‌های اندازه‌گیری ترکیب شدند.

در هر مرحله از دلفی، نتایج مرحله قبلی در یک گزارش بازخورد ارائه شد. از اعضای هیئت کارشناسان خواسته شد تا موافقت یا مخالفت خود را با توجه به طرح‌های پیشنهادی ارزیابی کنند. موافقت بر روی مقیاس 5 نقطه‌ای ارزیابی شد (کاملاً مخالف، مخالف، نظری ندارم، موافق، کاملاً موافق). اعضای هیئت استدلال‌های خود را در خصوص گزینه‌های خود برای متقاعد کردن اعضای دیگر ارائه کردند. اجماع در خصوص یک مسئله زمانی حاصل شد که حداقل 67 درصد اعضای هیئت عبارت "موافق یا کاملاً موافق" را در مقیاس 5 نقطه‌ای نشان داد. در صورتی که توافق کمتر از 67 درصد حاصل می‌شد، ما مجدداً در مرحله بعدی این خواسته را مطرح می‌کردیم. وقتی که هیچ‌گونه اجماعی حاصل نشد، کمیته اجرایی، تصمیم نهایی را گرفت.

در صورت لزوم، ما از اعضای هیئت خواستیم تا روش‌های آماری مطلوب را برای هر تئوری اندازه‌گیری به طور جداگانه نشان دهند یعنی تئوری آزمون کلاسیک (CTT) یا تئوری پاسخ سؤال (IRT) و یا برای هر نوع امتیاز نظیر دو بخشی، اسمی، ترتیبی یا پیوسته.

## نتایج

### اعضای هیئت

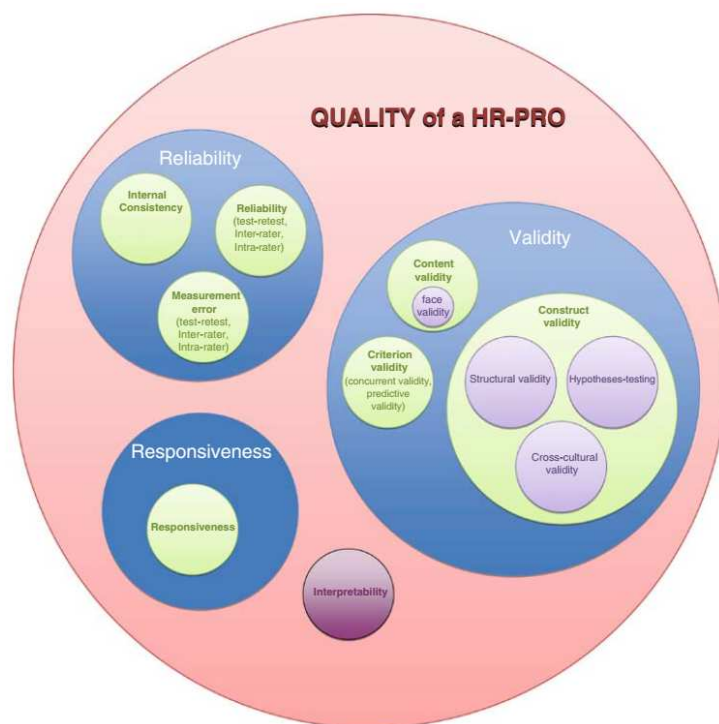
ما از 91 کارشناس برای شرکت دعوت به عمل آوردیم که 63 درصد (یا 57 نفر) آن‌ها شرکت کردند. دلیل اصلی عدم مشارکت، نبود زمان بود. 19 کارشناس (21 درصد) پاسخ ندادند. از 57 کارشناسی که شرکت کردند 43 نفر حداقل در یک مرحله شرکت کرده و 20 نفر (35 درصد) در همه چهار مرحله شرکت کردند. تعداد میانگین (حداقل - حداکثر) سال‌های سابقه کار در اندازه گیری سلامت یا رشته‌های مشابه (برای مثال در اندازه گیری‌های روان شناسی یا آموزشی)، 20 سال (6-40 به طور متوسط) بود. بیشتر اعضای هیئت مربوط به امریکای شمالی (25) و اروپا (29) بودند، در حالی که دو عضو مربوط به استرالیا و یک عضو متعلق به آسیا بود. نرخ پاسخ مراحل از 48 تا 74 درصد متغیر بود. شش عضو هیئت (11 درصد) در طی فرایند مطالعه را رها کردند. نام همه اعضای هیئت که حداقل یک بار شرکت کرده بودند در قسمت تشکر و قدردانی آورده شده است.

### طبقه بندی COSMIN

در مطالعه دلفی، ما یک طبقه بندی از روابط ویژگی‌های اندازه گیری که برای ارزیابی ابزارهای HR-PRO مناسب بودند را توسعه داده و در خصوص اصطلاحات و تعاریف این ویژگی‌های اندازه گیری به توافق رسیدیم. روابط بین همه ویژگی‌ها در طبقه بندی شکل 1 نشان داده شده است. طبقه بندی شامل سه حوزه (پایایی، روایی و پاسخگویی) می‌باشد که در بر گیرنده ویژگی‌های اندازه گیری است. ویژگی اندازه گیری (روایی سازه) شامل سه بعد است: روایی ساختاری، آزمون فرضیات و روایی بین فرهنگی. تفسیر پذیری در طبقه بندی و چک لیست قرار گرفت، اگرچه به عنوان یک ویژگی اندازه گیری در نظر گرفته نشده است ولی یک خصوصیت مهم است. درصد اجماع در خصوص اصطلاحات و جایگاه آن‌ها در طبقه بندی در بخش‌های دیگر توصیف شده است (12).

شکل 1: طبقه بندی COSMIN روابط ویژگی‌های اندازه گیری





| پایایی   | روایی  | پاسخگویی     |
|--|--|--------------|
| همسانی درونی   | روایی محتوایی (روایی صوری)                                     | پاسخگویی     |
| پایایی (آزمون، باز آزمون، بین ارزیاب، درون ارزیاب)           | روایی ملاکی (روایی همزمان، روایی پیش بین)                      | تفسیر پذیری  |
| خطای اندازه گیری (آزمون، باز آزمون، بین ارزیاب، درون ارزیاب) | روایی سازه<br>روایی ساختاری، آزمون فرضیات،<br>روایی بین فرهنگی | کیفیت HR-PRO |

### چک لیست COSMIN

نتایج اجماع حاصله در مراحل دلفی برای ایجاد چک لیست COSMIN (پیوست 2) استفاده شد. چک لیست دارای 12 باکس می‌باشد. ده باکس را می‌توان برای ارزیابی این که آیا یک مطالعه با معیار کیفیت روش شناسی خوب انطباق دارد یا خیر استفاده کرد. نه باکس حاوی استانداردها و معیارهایی برای ویژگی‌های اندازه گیری (همسانی درونی (باکس A)، پایایی (باکس B)، خطای اندازه گیری (باکس C)، روایی محتوایی (باکس D)، روایی

ساختاری (باکس E)، آزمون فرضیات (باکس F)، روایی بین فرهنگی (باکس G)، روایی ملاکی (باکس H) و پاسخگویی (باکس I) و یک باکس حاوی معیارهایی برای مطالعات در زمینه تفسیر پذیری (باکس J) است. به علاوه، دو باکس در چک لیست قرار دارند که در بر گیرنده ملزومات عمومی برای مقالاتی هستند که در آنها روش‌های IRT (باکس IRT) و ملزومات عمومی تعمیم پذیری نتایج (باکس تعمیم پذیری) قرار دارند.

به منظور تکمیل چک لیست COSMIN، یک روش چهار مرحله‌ای بایستی استفاده شود (شکل 2) (13). مرحله 1، تعیین این است که کدام ویژگی‌ها در یک مقاله ارزیابی شوند. مرحله 2، تعیین این است که آیا روش‌های آماری استفاده شده در مقاله بر اساس تئوری آزمون کلاسیک است یا بر اساس تئوری پاسخ سؤال. برای مطالعاتی که از IRT استفاده می‌کنند، باکس IRT بایستی کامل شود. مرحله سوم، تکمیل باکس‌ها با استانداردهایی است که با ویژگی‌های انتخاب شده در مرحله 1 همراه باشد. این باکس‌ها دارای سوالاتی برای ارزیابی این هستند که آیا یک مطالعه استانداردهای کیفیت روش شناسی خوب را دارد یا خیر. گویه‌ها در مورد ملزومات طراحی و روش‌های آماری مطلوب هر یک از ویژگی‌های اندازه گیری گنجانده می‌شوند. به علاوه، یک باکس با گویه‌هایی در خصوص تفسیر پذیری امتیاز تغییر گنجانده شده است (باکس J). تعداد گویه‌ها در این باکس‌ها از 5 تا 18 متغیر است. مرحله 4 این روش، تکمیل باکس مربوط به ملزومات عمومی تعمیم پذیری نتایج است. این باکس تعمیم پذیری بایستی برای هر ویژگی شناسایی شده در مرحله 1 تکمیل شود. ما یک راهنمایی را برای توصیف هر گویه و پیشنهادهای برای امتیاز دهی ایجاد کردیم (13).

### اجماع میان هیئت کارشناسان

در جدول 1، ما دامنه درصد توافق میان اعضای هیئت کارشناسان را برای هر باکس هم برای ملزومات طراحی و هم برای روش‌های آماری ارائه می‌کنیم. بیشتر این مسائل در مرحله 2 و 3 بحث شدند.

درصد توافق میان اعضا در رابطه با گویه‌های 1-3 در باکس IRT از 81 تا 96 درصد متغیر بود. گویه 4 (یعنی بررسی فرضیات برآورد پارامترهای مدل IRT) بر اساس پیشنهاد یکی از اعضا در مرحله 4 گنجانده شد. به این ترتیب هیچ اجماعی ارزیابی نشده و کمیته اجرایی تصمیم به گنجاندن این گویه گرفت.

چهار گویه لحاظ شده در چک لیست دارای توافق کم‌تر از 67 درصد بودند: گویه 9 از باکس A (همسانی درونی)، گویه 11 از باکس C (خطای اندازه‌گیری)، گویه‌های 11 و 17 از باکس A (پاسخگویی). همه به جز یک مورد مربوط به روش‌های آماری بودند. به دلایل متفاوت، کمیته اجرایی تصمیم به لحاظ کردن این چهار گویه در چک لیست گرفت.

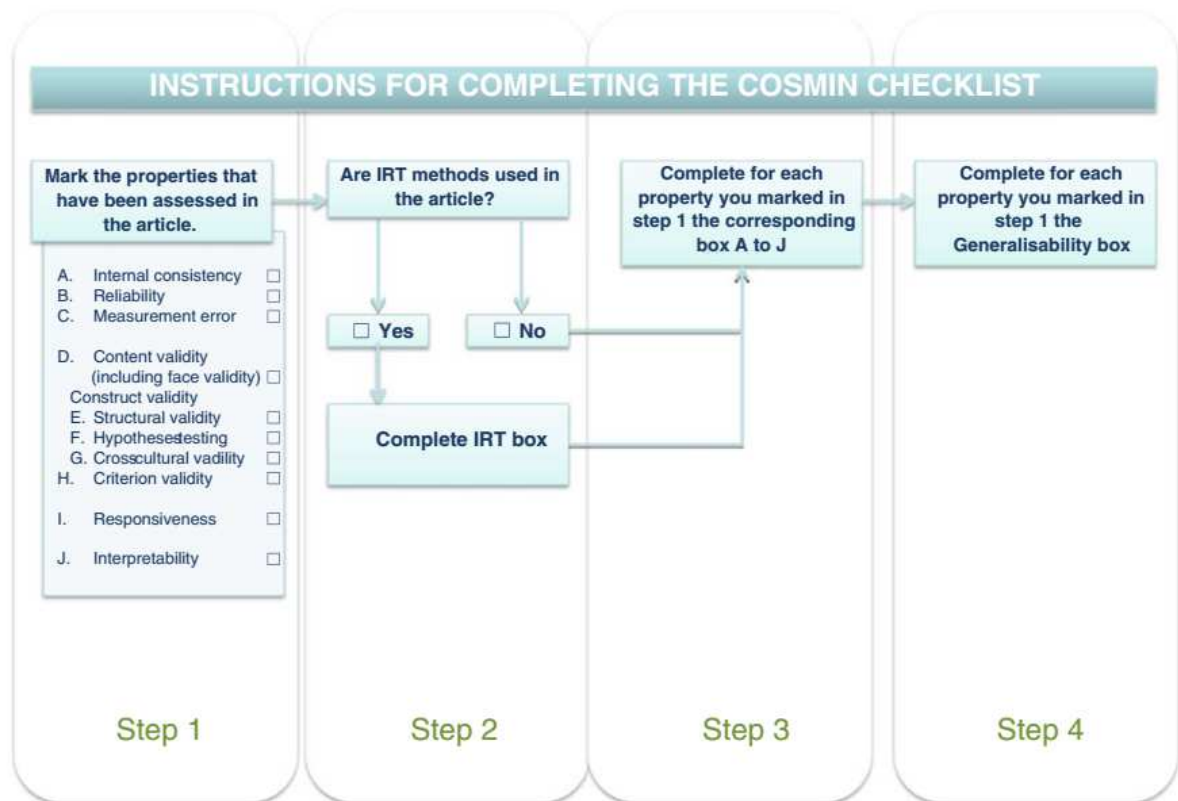
ما در هنگام پرسش در مورد روش آماری مطلوب برای همسانی درونی، بین انواع امتیازات یعنی دو بخشی یا ترتیبی (گویه 9) تمایزی قائل نشدیم. از این روی الفای کرونباخ نسبت به KR-20 مطلوب‌تر بود. با این حال، کمیته اجرایی تصمیم گرفت که KR-20 برای امتیازات دو بخشی نیز مناسب است.

گویه 11 از باکس C (خطای اندازه‌گیری) دارای سه روش یعنی خطای معیار اندازه‌گیری (SEM)، کوچک‌ترین تغییر قابل مشاهده (Sdc) و حدود توافق (LOA) بود. در مرحله 3، SEM به عنوان روشی مطلوب برای اندازه‌گیری خطای معیار انتخاب شد. هنگام پرسش در مورد سایر روش‌های مناسب، تنها 20 درصد با SDC و 28 درصد با LOA موافق بودند. علی‌رغم درصد توافق پایین حاصله در مرحله 3 در خصوص پذیرش SDC و LOA به عنوان روش‌های مناسب، کمیته اجرایی تصمیم گرفت که هر دو روش‌ها بایستی برای اندازه‌گیری خطای اندازه‌گیری مناسب هستند و از این روی در چک لیست لحاظ شدند. SDC تبدیل خطی SEM می‌باشد (14) یعنی  $1.96 \times \sqrt{2} \times SEM$ . چون SEM یک روش مناسب است، SDC نیز می‌تواند مناسب باشد. LOA پارامتری است که نشان می‌دهد میزان تفاوت دو شاخص چقدر است (15). وقتی که این دو شاخص شاخص‌های تکراری در بیماران باشند، می‌توان از آن به عنوان روشی برای ارزیابی خطای اندازه‌گیری استفاده کرد. LOA مستقیماً مربوط به SEM (16) بوده و ما تصمیم گرفتیم تا این روش را در چک لیست قرار دهیم.

گویه 11 از باکس A (پاسخگویی) یک توصیف کافی در خصوص ابزارهای مقایسه‌گر بود که مورد توافق 64 درصد اعضا قرار گرفت. اگرچه درصد توافق نسبتاً پایین بود، ما تصمیم گرفتیم تا این گویه را لحاظ کنیم زیرا در باکس F (آزمون فرضیات) قرار داشته و منعکس‌کننده تشابه بین روایی ساختار و پاسخگویی است.

گویه 17 از باکس A در برگیرنده دو روش یعنی همبستگی‌های بین امتیازات تغییر و سطح زیر منحنی می‌باشد. هفتاد و شش درصد هیئت کارشناسان، اولین روش را به صورت روش مطلوب در نظر گرفتند. این روش را زمانی

می‌توان استفاده کرد که هر دو ابزار اندازه‌گیری مورد مطالعه و استاندارد طلایی، شاخص‌های پیوسته‌ای هستند. تنها 60 درصد روش ROC را یک روش مناسب برای اندازه‌گیری پاسخگویی، در زمانی که استاندارد طلایی موجود بود در نظر گرفتند. کمیته اجرایی، روش ROC را یک روش مناسب برای ارزیابی در نظر گرفت به خصوص اگر یک ابزار اندازه‌گیری به اندازه استاندارد طلایی خوب باشد. از این روی کمیته اجرایی تصمیم به لحاظ کردن این تصمیم گرفت.



| دستور العمل‌هایی برای تکمیل چک لیست COSMIN   |   |   |   |
|--|---|---|---|
| ویژگی‌های ارزیابی شده در مقاله کدامند؟   | آیا روش‌های IRT در مقاله استفاده می‌شوند؟ | برای هر ویژگی تعیین شده در مرحله اول، باکس A تا J مربوطه را کامل کنید | برای هر ویژگی تعیین شده در مرحله اول، باکس تعمیم‌پذیری را کامل کنید |
| الف: همسانی درونی<br>ب: پایایی<br>پ: خطای اندازه‌گیری<br>ت: روایی محتوی (روایی صوری) | بله، خیر<br>باکس IRT را کامل کنید         |   |   |

|         |         |         |   |
|---------|---------|---------|---|
|         |         |         | ث: روایی ساختاری<br>ج: آزمون فرضیه<br>چ: روایی بین فرهنگی<br>خ: روایی ملاکی |
| مرحله 4 | مرحله 3 | مرحله 2 | مرحله 1   |

شکل 2: روش چهار مرحله‌ای برای تکمیل چک لیست COSMIN

جدول 1: درصد توافق اعضای هیئت کارشناسانی که با گویه‌های مربوط به نیازهای طراحی و روش‌های آماری

برای باکس‌های COSMIN A-J

| روش‌های آماری (درصد) | ملزومات طراحی (درصد) |  |
|----------------------|----------------------|--|
| 40-88 (R2-4)         | 77-92 (R2)           | همسانی داخلی                           |
| 80-92 (R3)           | 77-97 (R2)           | پایایی                                 |
| 20-76 (R3)           | گویه‌های مشابه برای  | خطای اندازه‌گیری                       |
| na                   | پایایی               | روایی محتوایی                          |
| 68-100 (R3)          | 90-94 (R2)           | روایی ساختاری                          |
| 90 (R2)              | 72 (R3)              | آزمون فرضیه                            |
| 68-94 (R3)           | 77-92 (R2, R4)       | روایی بین فرهنگی                       |
| 88 (R3)              | 70-79 (R3-4)         | روایی ملاکی                            |
| na                   | 88 (R3)              | پاسخگویی (عمومی)                       |
| 88 (R3)              | 90-97 (R2)           | پاسخگویی (عدم وجود استانداردهای طلایی) |
| 60-76 (R3)           | 64-68 (R3)           | پاسخگویی (وجود استانداردهای طلایی)     |
| 72-96 (R3)           | 80 (R3)              | تفسیر پذیری                            |
|                      | na                   |  |

در مطالعه دلفی، ما یک چک لیست حاوی استاندارد را برای ارزیابی کیفیت روش شناختی مطالعات در زمینه ویژگی‌های اندازه‌گیری توسعه دادیم. ما آن را برای تفکیک ارزیابی کیفیت روش شناختی مطالعه و ارزیابی نتایج آن مفید در نظر گرفتیم. چک لیست COSMIN ابزاری برای ارزیابی کیفیت روش شناختی مطالعه مربوط به ویژگی‌های اندازه‌گیری ابزار HR-PRO و نه برای ارزیابی کیفیت HR-PRO می‌باشد. برای ارزیابی کیفیت ابزار،

معیارهای ویژگی‌های اندازه‌گیری خوب بایستی قابل تعمیم به نتایج مطالعه در زمینه ویژگی‌های اندازه‌گیری باشد. مثال‌هایی از این معیارها توسط اعضای گروه ما (6) منتشر شده‌اند. با این حال، این معیارها مبتنی بر اجماع نبودند. توجه کنید که چک لیست COSMIN شامل این معیارهای کفایت نمی‌باشند.

اگرچه هدف اولیه ما توسعه این معیارها بود (17)، به دلیل کمبود زمان و پیچیدگی مسائل، ما معیارهای کفایت ویژگی‌های اندازه‌گیری را توسعه ندادیم. اجماع مربوط به این معیارها بایستی در آینده حاصل شود. به علاوه، توسعه یک سیستم ارزیابی و رتبه‌بندی که از طریق آن یک مطالعه را می‌توان به سطوح کیفیتی متفاوت برای مثال کیفیت روش شناختی عالی/خوب/نسبتاً خوب/ضعیف طبقه‌بندی کرد، مفید است.

چک لیست COSMIN را می‌توان برای ارزیابی کیفیت روش شناختی مطالعات مربوط به ویژگی‌های اندازه‌گیری ابزارهای اندازه‌گیری وضعیت سلامت استفاده کرد. برای مثال، این را می‌توان برای ارزیابی کیفیت یک مطالعه بر روی یک ابزار اندازه‌گیری و یا مقایسه ویژگی‌های اندازه‌گیری تعدادی از ابزارها استفاده کرد (18-19). در این ارزیابی، در نظر گرفتن کیفیت روش شناختی مطالعات منتخب مهم است. اگر نتایج مطالعات با کیفیت بالا متفاوت از نتایج مطالعات با کیفیت پایین باشد، این نشانه وجود اریبی است. چک لیست COSMIN را می‌توان به عنوان راهنمایی برای طراحی یا گزارش یک مطالعه در زمینه ویژگی‌های اندازه‌گیری استفاده کرد. به علاوه، دانشجویان در زمان یادگیری در مورد ویژگی‌های اندازه‌گیری می‌توانند از آن استفاده کنند و داوران و سردبیران مجلات می‌توانند از آن برای ارزیابی کیفیت روش شناختی مقالات استفاده کنند.

بحث‌های نظری‌ای وجود دارند که نشان می‌دهند نیاز مبرمی به یک ابزار برای اثبات پایایی، روایی و پاسخگویی خوب احساس می‌شود. تا آن جا که می‌دانیم، مارشال (3)، تنها کسی بود که به طور تجربی نشان داد که نتایج مطالعات در زمانی که ابزارهای اندازه‌گیری معتبر استفاده می‌شوند متفاوت می‌باشند. با این حال تحقیقات تجربی بیشتری بایستی برای تأیید این مسئله انجام شوند. مطالعات را می‌توان برای این منظور انجام داد که در آن نتایج کارآزمایی‌های کنترل شده تصادفی که از ابزارهای اندازه‌گیری مناسب استفاده می‌کند و RCT که از ابزارهای با پاسخگویی مجهول استفاده می‌کند مقایسه می‌شوند.

یک رویکرد دلفی برای شرایطی مفید است که در آن کمبود شواهد تجربی و تفاوت نظر زیادی وجود دارد. پاسخ‌های مربوط به سؤالات تحقیق مطالعه COSMIN به طور تجربی نمی‌تواند بررسی شود. از این روی، توافق میان کارشناسان مفید است. در منابع و مطالعات، مقدار معیار بین 55 و 100 درصد استفاده می‌شود (20). مقدار معیار 67 درصد برای اجماع به طور دلخواه استفاده شد.

انتخاب نمونه‌های تصادفی از میان همه کارشناسان غیرممکن است. از این روی انتخاب کارشناسان لزوماً غیر سیستماتیک بود. همه نویسندگان اول و آخر شناسایی شده توسط دو مرور منابع سیستماتیک توصیف شده در بخش روش‌ها به صورت کارشناسان بالقوه انتخاب شدند. ما افرادی را که کارشناس بودند را به لیست افزودیم. برای کنترل یک کارشناس، ما PubMed را چک کردیم تا ببینیم که آیا نویسنده حداقل 5 مقاله را در زمینه مسائل اندازه‌گیری منتشر کرده است یا خیر. مجموع 30 کارشناس به صورت کافی در نظر گرفته شد به طوری که تعداد زیادی از آن‌ها قابل مدیریت نبودند.

در این مطالعه، ما بر ابزارهای HR-PRO متمرکز بوده‌ایم. با این حال، ویژگی‌های اندازه‌گیری مشابه برای سایر انواع ابزارهای اندازه‌گیری مرتبط با سلامت نظیر ابزارهای مبتنی بر عملکرد و مقیاس‌های رتبه بندی بالینی وجود دارند. به علاوه، ما بر ابزارهای ارزیابی تمرکز کردیم. با این حال، برای اهداف پیش‌بینی یا تفکیکی، ملزومات طراحی و استانداردهای ویژگی‌های اندازه‌گیری مشابه می‌باشند.

چک لیست COSMIN، توصیه‌های عمومی را در خصوص اندازه‌گیری‌های HR-PRO ارائه می‌کند. برخی از استانداردها در چک لیست COSMIN نیازمند اصلاحات بیشتری می‌باشند برای مثال از طریق تعریف این که اندازه نمونه کافی چقدر است و یا از طریق بازه زمانی آزمون-باز آزمون کافی یا زمانی که چیزی به طور کافی توصیف می‌شود. چون این مسائل به شدت وابسته به ساختار اندازه‌گیری شده هستند، کاربران بایستی این تصمیمات را برای خود بگیرند.

برای کمک به کاربران آینده چک لیست COSMIN، ما اقدام به توصیف برخی بحث‌های موجود در مراحل دلفی در مورد معیارها (21) کردیم. در راهنما (13)، ما به توضیح هر آیت‌م پرداخته و پیشنهادهای را برای امتیاز بندی گویه‌ها در چک لیست ارائه می‌کنیم.

هدف رویکرد COSMIN، بهبود انتخاب ابزارهای اندازه گیری است. در مرحله اول، ما به یک اجماع رسیدیم که ویژگی‌های اندازه گیری مهم هستند. و ما استانداردهایی را برای شیوه ارزیابی این ویژگی‌های اندازه گیری توسعه دادیم. چک لیست COSMIN با مشارکت بسیاری از کارشناسان در رشته مربوطه ایجاد شد. این چک لیست موجب تسهیل انتخاب مناسب‌ترین شاخص HR-PRO در میان ابزارهای مختلف می‌شود. با مشارکت بسیاری از کارشناسان در فرایند توسعه چک لیست COSMIN، احتمال دارد که همه گویه‌های مربوطه از همه ویژگی‌های اندازه گیری مربوطه گنجانده شوند و این به تقویت روایی محتوی کمک می‌کند. به علاوه، هدف ما ارزیابی پایایی بین ارزیابی چک لیست COSMIN در گروه بزرگی از محققان بین المللی است.

## پیوست 2: چک لیست COSMIN

### مرحله 1: ویژگی‌های اندازه گیری ارزیابی شده در این مقاله

|  |                  |
|--|------------------|
|  | همسانی درونی     |
|  | پایایی           |
|  | خطای اندازه گیری |
|  | روایی محتوی      |
|  | روایی ساختاری    |
|  | آزمون فرضیه      |
|  | روایی بین فرهنگی |
|  | روایی ملاکی      |
|  | پاسخگویی         |
|  | تفسیر پذیری      |

مرحله 2: تعیین روش آماری مورد استفاده در مقاله بر اساس CTT یا IRT است

| باکس ملزومات عمومی برای مطالعاتی که از مدل‌های تئوری پاسخ گویه استفاده کرده‌اند |     |   |
|---|-----|---|
| خیر   | بله | 1- آیا مدل IRT به طور کافی توصیف شده است؟ برای مثال، مدل لجستیک یک پارامتری (OPLM)، مدل اعتبار جزیی (PCM)، مدل پاسخ درجه بندی شده (GRM) |
|   |     | 2- آیا نرم افزار کامپیوتری به طور کافی توصیف شده است؟ برای مثال   |



|  |  |   |
|--|--|---|
|  |  | 3- آیا روش برآورد مورد استفاده به طور کافی توصیف شده است؟ برای مثال، حداکثر درست نمایی شرطی (CML)، حداکثر درست‌نمایی حاشیه‌ای (MML) |
|  |  | 4- آیا فرضیات برآورد پارامترهای مدل IRT چک شده‌اند یا خیر؟ برای مثال، تک بعدی بودن، استقلال محلی و برازش گویه (IDF)                 |
|  |  |   |

مرحله 3: تعیین این که آیا مطالعه مطابق با استانداردهای کیفیت روش شناسی خوب است؟

| باکس A: همسانی درونی |     |   |
|----------------------|-----|---|
| بله                  | خیر | 1- آیا مقیاس متشکل از شاخص‌های اثر می‌باشد یعنی آیا بر اساس یک مدل انعکاسی است؟ |
|                      |     | ملزومات طراحی   |
|                      |     | 2- آیا درصد گویه های مفقود مشخص است؟  |
|                      |     | 3- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟                                |
|                      |     | 4- آیا اندازه نمونه در تحلیل همسانی درونی کافی است؟                             |
|                      |     | 5- آیا تک بعدی بودن مقیاس چک شده است؟   |
|                      |     | 6- آیا اندازه نمونه در تحلیل تک بعدی کافی است؟                                  |
|                      |     | 7- آیا آماره همسانی درونی برای هر مقیاس به طور جداگانه محاسبه شده است؟          |
|                      |     | 8- آیا نقص‌های مهمی در طراحی یا روش‌های مطالعه وجود دارد؟                       |
|                      |     | روش‌های آماری   |
|                      |     | 9- آیا برای تئوری آزمون کلاسیک، الفای کرونباخ محاسبه شده است؟                   |
|                      |     | 10- آیا برای امتیازات دو بخشی، الفای کرونباخ محاسبه شده است یا KR-20؟           |
|                      |     | 11- آیا برای IRT، آماره خوبی برازش در سطح جهانی محاسبه شده است؟                 |
|                      |     | برای مثال کای اسکوئر، ضریب پایایی مقدار صفت پنهان برآورد شده                    |

| باکس B: پایایی: شاخص‌های نسبی (از جمله آزمون-بازآزمون، پایایی بین ارزیاب و پایایی درون ارزیاب |     |  |
|---|-----|--|
| بله   | خیر | ملزومات طراحی                            |
|   |     | 1- آیا درصد گویه های مفقود بیان شده است؟ |

|  |  |  |
|--|--|--|
|  |  | 2- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟   |
|  |  | 3- آیا اندازه نمونه در تحلیل، کافی است؟  |
|  |  | 4- آیا حداقل دو اندازه گیری انجام شده است؟   |
|  |  | 5- آیا اندازه گیری ها مستقل هستند؟   |
|  |  | 6- آیا بازه زمانی مشخص شده است؟  |
|  |  | 7- آیا بیماران در دوره موقت در ساختار اندازه گیری شده پایدار بوده اند؟                   |
|  |  | 8- آیا بازه زمانی مناسب بود؟   |
|  |  | 9- آیا شرایط آزمایش برای دو اندازه گیری مشابه است؟ برای مثال نوع پرسش نامه، محیط و آموزش |
|  |  | 10- آیا نقص های مهمی در طراحی و روش های مطالعه وجود دارد؟                                |
|  |  | <b>روش های آماری</b>   |
|  |  | 11- برای امتیازات ییوسته: آیا ضریب همبستگی بین دسته های محاسبه شده است؟                  |
|  |  | 12- آیا برای امتیازات ترتیبی/ دو بخشی، ضریب کاپا محاسبه شده است؟                         |
|  |  | 13- برای امتیازات ترتیبی، آیا ضریب کاپای وزنی محاسبه شده است؟                            |
|  |  | 14- آیا برای امتیازات ترتیبی، طرح وزن دهی توصیف شده است؟ برای مثال خطی، دو جمله ای       |

|   |     |  |
|---|-----|--|
| <b>باکس C: خطای اندازه گیری: شاخصهای مطلق</b> |     |  |
|   |     | <b>ملزومات اندازه گیری</b>                       |
| بله   | خیر | 1- آیا درصد گویه های مفقود بیان شده است؟         |
|   |     | 2- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟ |

|  |  |   |
|--|--|---|
|  |  | 3- آیا اندازه نمونه در تحلیل، کافی است؟   |
|  |  | 4- آیا حداقل دو اندازه گیری انجام شده است؟  |
|  |  | 5- آیا اندازه گیری‌ها مستقل هستند؟  |
|  |  | 6- آیا بازه زمانی مشخص شده است؟   |
|  |  | 7- آیا بیماران در دوره موقت در ساختار اندازه گیری شده پایدار بوده‌اند؟                                |
|  |  | 8- آیا بازه زمانی مناسب بود؟  |
|  |  | 9- آیا شرایط آزمایش برای دو اندازه گیری مشابه است؟ برای مثال نوع پرسش<br>نامه، محیط و آموزش           |
|  |  | 10- آیا نقص‌های مهمی در طراحی و روش‌های مطالعه وجود دارد؟   |
|  |  | روش‌های آماری   |
|  |  | 11- آیا برای CTT، خطای معیار اندازه گیری، کوچک‌ترین تغییر قابل<br>مشاهده و حدود توافق محاسبه شده است؟ |

| باکس D: روایی محتوایی (از جمله روایی صوری) |     |   |
|--|-----|---|
| بله  | خیر | ملزومات عمومی   |
|  |     | 1- آیا ارزیابی این که همه گویه‌ها اشاره به ابعاد مرتبط با ساختار مورد اندازه<br>هستند یا خیر صورت گرفته است؟                    |
|  |     | 2- آیا ارزیابی این که همه گویه‌ها برای جمعیت مورد مطالعه مناسب هستند<br>صورت گرفته است (برای مثال سن، جنس، ویژگی بیماری، کشور)؟ |
|  |     | 3- آیا ارزیابی این که همه گویه‌ها برای هدف ابزار اندازه گیری مناسب هستند<br>صورت گرفته است (افتراقی، ارزشگذارانه، پیش بین)؟     |
|  |     | 4- آیا ارزیابی این که همه گویه‌ها منعکس کننده ساختار مورد اندازه گیری   |

|  |  |   |
|--|--|---|
|  |  | هستند صورت گرفته است؟                               |
|  |  | 5- آیا نقص‌های مهمی در طرح یا روش مطالعه وجود دارد؟ |

| باکس E: روایی ساختاری |     |   |
|-----------------------|-----|---|
| خیر                   | بله | 1- آیا مقیاس متشکل از شاخص‌های اثر است یعنی بر اساس مدل انعکاسی است؟          |
|                       |     | ملزومات طراحی   |
|                       |     | 2- آیا درصد داده‌های مفقود ارائه شده است؟                                     |
|                       |     | 3- آیا شیوه مدیریت گویه‌های مفقود توصیف شده است؟                              |
|                       |     | 4- آیا اندازه نمونه کافی است؟   |
|                       |     | 5- آیا نقص‌های مهمی در طرح یا روش مطالعه موجود است؟                           |
|                       |     | روش‌های آماری   |
|                       |     | 6- آیا برای CTT، تحلیل عاملی تاییدی یا اکتشافی صورت گرفته است؟                |
|                       |     | 7- آیا برای IRT، تست‌های IRT برای تعیین تک بعدی بودن گویه‌ها، صورت گرفته است؟ |

| باکس F: آزمون فرضیه |     |   |
|---------------------|-----|---|
| خیر                 | بله | ملزومات طراحی   |
|                     |     | 1- آیا درصد گویه‌های مفقود مشخص است؟  |
|                     |     | 2- آیا شیوه مدیریت گویه‌های مفقود توصیف شده است؟  |
|                     |     | 3- آیا اندازه نمونه تحلیل کافی است؟   |
|                     |     | 4- آیا فرضیات مربوط به همبستگی‌ها یا تفاوت میانگین‌ها قبلاً مشخص شده است (قبل از جمع‌آوری داده‌ها)؟ |

|  |  |
|--|--|
|  | <p>5- آیا جهت مورد انتظار همبستگی‌ها یا تفاوت میانگین‌ها در فرضیات مشخص است؟</p> <p>6- آیا بزرگی نسبی یا مطلق مورد انتظار همبستگی‌ها یا تفاوت‌های میانگین در فرضیه لحاظ شده است؟</p> <p>7- آیا برای روایی همگرایی، توصیف کافی در خصوص ابزارهای مقایسه گر ارائه شده است؟</p> <p>8- آیا برای روایی افتراقی، ویژگی‌های اندازه گیری ابزارهای مقایسه گر به طور کافی مقایسه شده است؟</p> <p>9- آیا نقص‌های مهمی در طرح یا روش‌های مورد مطالعه وجود دارد؟</p> <p style="text-align: right;"><b>روش‌های آماری</b></p> <p>10- آیا روش‌های آماری و طرح برای تست فرضیه‌ها کافی هستند؟</p> |
|--|--|

| <b>باکس G: روایی بین فرهنگی</b> |     |   |
|---------------------------------|-----|---|
| <b>ملزومات طراحی</b>            |     |   |
| خیر                             | بله |   |
|                                 |     | 1- آیا درصد گویه های مفقود مشخص است؟  |
|                                 |     | 2- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟  |
|                                 |     | 3- آیا اندازه نمونه تحلیل کافی است؟   |
|                                 |     | 4- آیا هر دو زبان اصلی که در آن ابزار HR-PRO توسعه یافته است و زبانی که در آن این ابزار ترجمه شده است، توصیف گردیده است؟  |
|                                 |     | 5- آیا تخصص افراد شرکت کننده در فرایند ترجمه به خوبی توصیف شده است؟ برای مثال تخصص در زمینه بیماری مربوطه، تخصص در زمینه ساختار اندازه گیری شده و تخصص در هر دو زبان؟ |
|                                 |     | 6- آیا مترجمان مستقل از هم کار کرده‌اند؟  |
|                                 |     | 7- آیا گویه ها به صورت پیشرو ترجمه شده‌اند یا پسرو؟   |
|                                 |     | 8- آیا یک توصیف کافی از شیوه حل تفاوت‌های بین نسخه‌های اصلی و ترجمه شده ارائه شده است؟  |
|                                 |     | 9- آیا ترجمه توسط کمیته ارزیابی شده است؟  |

|  |  |
|--|--|
|  | <p>10- آیا ابزار HR-PRO (مصاحبه‌های شناختی) برای کنترل تفسیر، اهمیت فرهنگی ترجمه و سهولت درک، پیش آزمون شده است؟</p> <p>11- آیا نمونه مورد استفاده در پیش آزمون به طور کافی توصیف شده است؟</p> <p>12- آیا نمونه‌ها از نظر همه ویژگی‌ها به جز زبان یا پیش زمینه فرهنگی، مشابه بوده‌اند؟</p> <p>13- آیا نقص‌های مهمی در طرح یا روش‌های مورد مطالعه وجود دارد؟</p> <p>روش‌های آماری:</p> <p>14- آیا برای CCT، تحلیل عاملی تاییدی وجود داشت؟</p> <p>15- آیا برای IRT، یک DIF بین گروه‌های زبانی ارزیابی شده وجود دارد؟</p> |
|--|--|

| باکس H: روایی ملاکی |     |  |
|---------------------|-----|--|
| خیر                 | بله | ملزومات طراحی  |
|                     |     | 1- آیا درصد گویه های مفقود مشخص است؟                                       |
|                     |     | 2- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟                           |
|                     |     | 3- آیا اندازه نمونه تحلیل کافی است؟  |
|                     |     | 4- آیا معیار مورد استفاده، یک استاندارد طلایی منطقی است؟                   |
|                     |     | 5- آیا نقص‌های مهمی در طرح یا روش مطالعه وجود دارد؟                        |
|                     |     | روش‌های آماری  |
|                     |     | 6- آیا برای امتیازات پیوسته، همبستگی‌ها یا مساحت زیر منحنی محاسبه شده است؟ |
|                     |     | 7- آیا برای امتیازات دو بخشی، حساسیت و ویژگی تعیین شده است؟                |

| باکس L: پاسخگویی |     |               |
|------------------|-----|---------------|
| خیر              | بله | ملزومات طراحی |
|                  |     |               |

|  |  |  |
|--|--|--|
|  |  | 1- آیا درصد گویه های مفقود مشخص است؟   |
|  |  | 2- آیا شیوه مدیریت گویه های مفقود توصیف شده است؟   |
|  |  | 3- آیا اندازه نمونه تحلیل کافی است؟  |
|  |  | 4- آیا یک طرح طولی با حداقل دو اندازه گیری استفاده شده است؟  |
|  |  | 5- آیا بازه زمانی بیان شده است؟  |
|  |  | 6- آیا رخ داده در بازه زمانی موقت (مداخله، یا رویدادهای دیگر) به طور کافی توصیف شده است؟   |
|  |  | 7- آیا نسبت بیماران تغییر کرده است (پیشرفت یا پسرفت)؟<br>ملزومات طراحی برای آزمون فرضیه  |
|  |  | 8- آیا فرضیات مربوط به تغییرات در امتیازات از قبل تعیین شده اند (قبل از جمع اوری داده ها)؟   |
|  |  | 9- آیا جهت مورد انتظار همبستگی ها یا تفاوت های میانگین امتیازات تغییر ابزارهای HR-PRO در این فرضیات لحاظ شده است؟  |
|  |  | 10- آیا بزرگی نسبی یا مطلق همبستگی ها یا تفاوت میانگین امتیازات تغییر ابزارهای HR-PRO در فرضیات لحاظ شده است؟  |
|  |  | 11- آیا توصیف کافی در مورد ابزارهای مقایسه گر ارائه شده است؟   |
|  |  | 12- آیا ویژگی های اندازه گیری ابزارهای مقایسه گر به طور کافی توصیف شده است؟  |
|  |  | 13- آیا نقص های مهمی در طرح یا روش مطالعه وجود دارد؟<br><b>روش های آماری</b>   |
|  |  | 14- آیا روش های آماری و طرح برای تست فرضیات کافی هستند؟<br>ملزومات طراحی برای مقایسه با استاندارد طلایی<br>برای ساختارهایی که برای آن ها یک استاندارد طلایی موجود است: |

|  |  |   |
|--|--|---|
|  |  | 15- آیا معیار تغییر، یک استاندارد طلایی منطقی است؟  |
|  |  | 16- آیا نقص‌های مهمی در طرح یا روش مطالعه وجود دارد؟  |
|  |  | روش‌های آماری   |
|  |  | 17- آیا برای امتیازات پیوسته، همبستگی‌های بین امتیازات تغییر و یا سطح زیر منحنی محاسبه شده است؟ |
|  |  | 18- آیا برای مقیاس‌های دو بخشی، حساسیت و ویژگی تعیین شده است؟                                   |

| باکس ۱: تفسیر پذیری |     |   |
|---------------------|-----|---|
| خیر                 | بله | 1- آیا درصد گویه‌های مفقود مشخص است؟  |
|                     |     | 2- آیا شیوه مدیریت گویه‌های مفقود توصیف شده است؟  |
|                     |     | 3- آیا اندازه نمونه تحلیل کافی است؟   |
|                     |     | 4- آیا توزیع کل امتیازات در نمونه مطالعه توصیف شده است؟   |
|                     |     | 5- آیا درصد پاسخگویان با کم‌ترین امتیاز ممکن توصیف شده است؟   |
|                     |     | 6- آیا درصد پاسخگویان با بیشترین امتیاز ممکن توصیف شده است؟   |
|                     |     | 7- آیا امتیازات و امتیازات تغییر (میانگین و انحراف معیار) برای گروه‌ها ارائه شده است؟ برای مثال، برای گروه‌های هنجاری، زیر گروه بیماران یا جمعیت عمومی؟ |
|                     |     | 8- آیا تغییرات مهم حداقل (MIC) یا تفاوت‌های مهم حداقل (MID) تعیین شده است؟  |
|                     |     | 9- آیا نقص‌های مهمی در طرح یا روش مطالعه وجود دارد؟   |



مرحله 4: تعیین تعمیم پذیری نتایج

| باکس تعمیم پذیری |  |
|------------------|--|
|                  | آیا نمونه‌ای که در آن ابزار HR-PRO ارزیابی شده است از حیث موارد زیر به طور کافی توصیف شده است؟                               |
|                  | 1- سن میانه یا میانگین (با انحراف معیار یا دامنه)؟   |
|                  | 2- توزیع جنسی؟   |
|                  | 3- ویژگی‌های بیماری مهم (برای مثال شدت، وضعیت، مدت) و توصیف درمان؟   |
|                  | 4- شرایطی که در آن مطالعه انجام شده است؟ برای مثال جمعیت کل، مراقبت‌های اولیه یا مراقبت‌های توان بخشی و بستری                |
|                  | 5- کشورهایی که در آن مطالعه انجام شده است؟   |
|                  | 6- زبانی که در آن ابزار HR-PRO ارزیابی شده است   |
|                  | 7- آیا روش مورد استفاده برای انتخاب بیماران به طور کافی توصیف شده است؟ برای مثال روش نمونه برداری در دسترس، متوالی یا تصادفی |
|                  | 8- آیا درصد پاسخ‌های مفقود (نرخ پاسخ) قابل قبول است؟   |



این مقاله، از سری مقالات ترجمه شده رایگان سایت ترجمه فا میباشد که با فرمت PDF در اختیار شما عزیزان قرار گرفته است. در صورت تمایل میتوانید با کلیک بر روی دکمه های زیر از سایر مقالات نیز استفاده نمایید:

لیست مقالات ترجمه شده ✓

لیست مقالات ترجمه شده رایگان ✓

لیست جدیدترین مقالات انگلیسی ISI ✓

سایت ترجمه فا ؛ مرجع جدیدترین مقالات ترجمه شده از نشریات معتبر خارجی