

پژوهش در جراحی

مثال‌های جراحی در سیستم‌های امتیازبندی

ترجمه و تلخیص: دکتر سیدعباس میرمالک* و گروه مترجمین**

مقدمه

آیا سیستم‌های امتیازبندی و مقیاس‌بندی، شعری علمی هستند یا ضرورتی بالینی؟ برخی از مثال‌های رایج جراحی ما را در پاسخگویی به این سوال یاری می‌دهند. در یک بیمار ۶۰ ساله پس از انجام همی کولتکومی سمت راست، مرحله‌بندی Duke به عنوان ابزاری توصیفی در برنامه‌ریزی درمانی آینده، در سطح وسیعی پذیرفته شده است. شیمی‌درمانی همراه بعد از عمل، امروزه روش درمانی توصیه شده‌ای برای سرطان روده بزرگ در Dukes C که برداشته شده است، می‌باشد. این مثالی در یک سیستم مرحله‌بندی ساخته شده، براساس پاتولوژی سرطان روده بزرگ است و هم اکنون استلزام درمانی دارد.

در زنی ۶۰ ساله با علائم پایدار سوزش قلب و رگورژیناسیون احتیاج به ارزیابی ۲۴ ساعته قسمت پائین مری برای تشخیص قطعی بیماری ریفلاکس گاستروازوفازیال دارد. تفسیر تصویر خام PH مری در ۲۴ ساعت موضوعی قابل بحث می‌باشد. روش علمی برای تمایز بین نوع فیزیولوژیک و شکل پاتولوژیک در معرض اسید قرار گرفتن بخش پائینی مری، استفاده از یک سیستم امتیازبندی است که برخی معیارهای مربوط به تصویر PH را در بر می‌گیرد. یعنی درصد زمانی که PH کمتر از ۴ است و تعداد دفعات و طول مدت ریفلاکس در یک سیستم امتیازبندی که بین در معرض اسید قرار گرفتن، ریفلاکس پاتولوژیک و طبیعی تمایز قائل می‌شود. چنین سیستمی برای متخصصین گوارش شناخته شده و قابل دسترسی است. این روش می‌تواند برای تعریف بیماری ریفلاکس گاستروازوفازیال مورد استفاده قرار گیرد. این مثالی از یک سیستم امتیازبندی علمی و ضروری با درگیری‌های تشخیصی و درمانی است که به عنوان یک استاندارد مرجع به شکل فزاینده‌ای مورد قبول واقع شده است، اگرچه هنوز به سطح مقبولیت طبقه‌بندی Duke که مثال زده شد، نرسیده است.

در مثالی دیگر زنی ۲۰ ساله با درد در قسمت ۱/۴ تحتانی راست شکم و استفراغ در طول یک روز به نظر می‌رسد که آپاندیسیت یا عفونت لگنی داشته باشد. پس از رد بیماری‌های التهابی لگن، جراحان ماهر و با تجربه آپاندیسیت را براساس شرح حال و یافته‌های بالینی و سونوگرافی تشخیص می‌دهند. هر چند که فهرستی طولانی از امتیازهای آپاندیسیت در دسترس است، لیکن کاربرد بالینی آنها اندک و قابل چشم‌پوشی است. امتیازهای پیچیده ارزش پیش‌بینی کننده نداشته و به شکل معتبری مورد استفاده قرار نمی‌گیرند.

نویسنده پاسخگو: دکتر سیدعباس میرمالک

تلفن: ۸۸۷۸۷۵۶۱

Email: SAM@Mirmalek.net

* استادیار گروه جراحی عمومی، دانشگاه علوم پزشکی آزاد اسلامی، واحد تهران

** دکتر لیلا پرورش، دکتر شقایق تهرانی، دکتر پوریا حسینی، دکتر پانته‌آ رضائیان، دکتر مریم سعیدیان،

دکتر مروا طهماسبی‌راد، دکتر علی غلامرضائزاد، دکتر مهدی کلاتری، دکتر الهام کنی، دکتر امیر تیمور مرعشی،

دکتر امید میرمطلبی، دکتر علی ناظمیان

تاریخ وصول: ۱۳۹۴/۱۰/۰۱

تعاریف

نیاز به تخصیص اعداد به پدیده‌های مشخص یا کیفیت، ضرورتی ساده است که به شدت قابل بحث می‌باشد. به ندرت سؤالی وجود دارد که در آن ضرورت محاسبه تعداد بیماران یا اندازه‌گیری سن به وسیله شمردن سال‌های زندگی مطرح باشد. براساس گفته Postman در کتاب Technopoly، در ۲۰۰ سال قبل کسی به بررسی یک نوشته به صورت تعیین کمیت آن یا تفهیم هوش یک فرد به وسیله اعداد واقعی نیندیشیده بود. امروزه امتیازهای مربوط به هوش یا یک درجه در یک نوشته، ابزارهای عددی تثبیت شده‌ای هستند و در جهت تعیین ارزش‌های کمیته‌اندیشه انسانی، قابل قبولند.

ما پذیرفته‌ایم که هرچیزی برای این که جنبه علمی یافته و قابل اعمال مدیریت باشد، باید به محاسبه درآید و اندازه‌گیری شود.

چه چیز باید اندازه‌گیری شود؟ احساس قابل اندازه‌گیری نیست. هیچ کس تاکنون سعی نکرده است که عشق و محبت را اندازه‌گیری کند با این همه، امتیازبندی، مقیاس‌بندی و مرحله‌بندی روندهای ریاضی هستند که در زندگی روزمره و طبابت بالینی به کار می‌روند. مقیاس‌ها ممکن است بسیار ساده یا پیچیده باشد. آنها ممکن است، شامل یک تصمیم ساده دوگانه بله یا خیر شوند، یا نظم درجه‌بندی که از اعداد کوچک‌تر به سمت اعداد بزرگ‌تر برود یا این که متغیر مداومی از صفر تا بی‌نهایت باشد. تمایز بین ۴ روش استفاده از اعداد در کلیه روش‌های مقیاس‌بندی بنیادی است و در سلسله مراتب ریاضی مطرح می‌شود. پایین‌ترین سطح طبقه‌بندی ساده و گروه‌بندی افراد یا موارد در مقوله‌های مختلف مانند: «مرد» یا «زن» می‌باشد. در نوع دوم، اعداد ترتیبی تخصیص داده می‌شوند و این اعداد بازتابی از نظر رو به افزایش متغیرهایی است، که باید اندازه‌گیری شوند. ارزش واقعی اعداد و فاصله بین آنها هنوز معنای درونی ندارد. در سومین نوع مقیاس فاصله معینی بین مقوله‌ها ایجاد می‌شود. این به آن معنی است که در مقیاس‌های فاصله‌ای، تغییر یک واحد تغییر مداوم مجموعه طیف مقیاس را ارائه می‌کند. تغییرات درجه حرارت مثال خوبی است. در «امتیازهای فاصله‌ای» اختلافات بین امتیازها قابل محاسبه می‌باشد. زیرا جمع و تفریق امتیازها مجاز است. در هر حال، مجاز به بیان این که یک امتیاز چند برابر امتیاز دیگر می‌باشد، نیستیم و این امر تنها در «امتیازهای نسبی» که شامل نقطه صفر است، مجاز شمرده می‌شود.

مقیاس‌ها و امتیازها نشان‌دهنده ساختارهای دشوارتری مانند ناتوانی و فعالیت‌های زندگی روزمره یا کیفیت زیست است که می‌تواند ترکیبی از اجزاء متفاوت باشد. به طور مثال، مقیاس‌های کیفیت زیست شامل علائم، عواطف و عملکردهای اجتماعی و جسمانی می‌باشد. تعریف مرحله‌بندی بیان می‌کند که یک بیماری در بیماران مختلف درجات مختلفی از شدت دارد یا این که بیماری با زمان تغییر می‌کند.

تعریف مرحله‌بندی تومور عنوان درجات شدت بیماری در بیماران مختلف است که در طی زمان تغییر می‌کند. به طور مثال، رشد منطقه‌ای تومور (گروه T)، درگیری غدد لنفاوی با سلول‌های بدخیم (گروه N) و متاستاز دوردست (گروه M)، به شکل منطقی انتشار منطقه‌ای تومور اپی‌تلیال را توصیف می‌کند (سرطان معده). به هر حال، سن، جنسیت و قرارگیری یک تومور در مرکز یا دوردست ممکن است در پیش‌آگهی و نتیجه یک بیماری نقش مهمی بازی کند. علاوه بر این، سیستم مرحله‌بندی می‌بایست نتایج درمانی یا تشخیصی داشته باشد. یا این که نتیجه یا پیش‌آگهی‌های مختلفی را تعریف کند. سیستم‌های مرحله‌بندی بدون نتیجه‌ای خاص، تنها «اشعار علمی» هستند.

معنی و کاربرد کلمه امتیازبندی "Scoring" دارای دو وجه است. یک سیستم امتیازبندی روشی است که اطلاعات بالینی را به منظور اتخاذ یک تصمیم به صورت اعداد ریاضی فشرده می‌کند. براساس نقطه مشخصی که به شکل منطقی انتخاب شده، یک تصمیم بالینی (بله / خیر یا طبیعی / غیرطبیعی) گرفته می‌شود. انتخاب بستگی به حساسیت یا تخصیص سیستم امتیازبندی دارد. در یک سیستم فرضی امتیازبندی، ارزش ویژه یک بیمار امتیاز نیز خوانده می‌شود. به طور مثال، یک بیمار ممکن است تحت عنوان یک امتیاز مشخص در مقیاس کوما‌ی گلاسکو محاسبه شود.

بر اساس توصیف دوبونو (De Bono) از تفکر انسانی، مغز انسان ادراک را از بیرون گرفته به صورت الگوهای مشخص سازماندهی می‌کند. شناخت الگوها و مقایسه نمونه‌های جاندار، کلید پیشرفت تفکر انسانی است. در مواجهه با هر پدیده‌ای، مغز این پدیده‌ها را در گروه‌های مشابهی که قبلاً با آنها مواجه شده است، جمع‌آوری کرده و در نتیجه می‌تواند به تناسب بر روی آنها عمل کند. سازوکار دقیق این شناخت نمونه‌ها در سطحی گسترده در ضمیر ناخودآگاه باقی می‌ماند. برای مثال بسیاری از مردم می‌دانند که چگونه یک تلویزیون یا دستگاه ضبط و پخش ویدیو را به راه بیاندازند، هر چند که از سازوکار دقیق دریافت و ایجاد تصویر تلویزیون آگاهی دقیقی ندارند. دوبونو این پدیده را سازوکار «جعبه سیاه» نامیده است. به عبارت دیگر، مغز ادراکات را ساده و فشرده کرده و آنها را در جعبه سیاه قرار می‌دهد و برچسب تلویزیون یا «تشخیص بیماری ریفلکس» بر آن می‌زند. این روند، ارتباط بین افراد را امکان‌پذیر می‌سازد. هدف سیستم‌های امتیازبندی نیز دقیقاً همین است: این سیستم‌ها اطلاعات بیولوژیک فردی و پیچیده را به صورت یک عدد واحد یا یک ردیف از حروف یا نمادها فشرده می‌کنند. باید بخاطر داشت که این تغییر، فشردگی و تعویض پایه واقعی اطلاعات است که موجب ارائه غلط یا سادگی بیش از حد اطلاعات اصلی می‌شود. این روش پردازش اطلاعات، فردیت را برمی‌اندازد و کاهش دهنده‌ای است که تمامیت و فردیت را در برخورد با بیمار، به فراموشی می‌سپارد. سیستم امتیازبندی می‌تواند به سه روش مختلف تشکیل شود: بر اساس امکانات بالینی یا اهمیت داده‌های آماری و یا بر اساس تلفیقی از هر دو روش. امکانات بالینی آشکار هستند. آلبومین، بیلی‌روبین، آسیت و کارکرد عصبی متغیرهایی منطقی برای ارزیابی عملکرد کبد هستند. رشد تومور، غدد لنفاوی یا متاستاز دیررس معیارهای محتمل برای ارزیابی رشد و پیشرفت تومور هستند. رویکرد روش‌شناختی برای تشکیل، ارزیابی و استفاده از سیستم‌های امتیازبندی به وسیله چارلسون و همکارانش (Charlson) توصیف شده است. مشکلات عملی و راه‌حل‌های آن در تعیین شاخص کیفیت زیست بیماران پیوند کبدی به وسیله پل و بویلون (Bouillon, Paul) توضیح داده شده است.

کاربردهای عملی برای سیستم‌های امتیازبندی

بر اساس بسیاری از امتیازات قابل دسترسی در نوشته‌ها، امتیازی که به طور همگانی شناخته شده و مورد قبول باشد ممکن است با یک محصول موفق صنعتی که به وسیله اسم شرکت و نه اسم اصلی محصول شناخته شده و مقایسه می‌شود. برای مثال Kleenex یک تکه برای خشک کردن چیزهای مختلف است. Xerox در انگلیسی آمریکایی به معنای فتوکپی کردن چیزی است. به همین صورت Apache, Nyha, Aitken, Breslow, Forreot, Child, Apgar, Duke اسامی هستند که برای مرحله‌بندی و امتیازبندی به خوبی شناخته شده‌اند و احتیاجی نیست که به کاربرد اصلی آنها اشاره شود. این اسامی در طول زمان خود، توصیف‌کننده شده‌اند. برخی از مثال‌ها برای سیستم‌های امتیازبندی در جدول ۱ بر اساس میزان مقبولیت و کاربرد آنها تقسیم شده‌اند.

جدول ۱ - مثال هایی از سیستم امتیازبندی

منبع) هدف	امتیازبندی هایی که به خوبی شناخته شده و در سطح وسیعی پذیرفته شده اند
انتشار تومور و مرحله	TNM
Vital function in noenates	APGAR
شکستگی در بزرگسالی	Aitken
کوما	Glasgow Coma Scale
شدت آسیب دیدگی	Traum Ascore
موقعیت فیزیولوژیک و سن	APACHE Score
عملکرد کبد	Child Classification
سرطان کولون	Dukes Classification
کیفیت زیست	Spitzer Index
Unguina hernia type	Nyhus Hernia Classification
فعالیت های جسمانی	Activity of Daily Life (Katz)
Symptomatic well-being	Visick Scale
Esoph acid exposure	امتیازبندی های تقریباً شناخته شده و مفید
Dyspepsia	De Meester Reflux Score
Inflammatory actinty Peritonitis	Talley Dyspepsia Sore
شدت آسیب دیدگی	Best – Indoaz for Crohns Disease Mannheim Peritonitis Index
بیماری و سلامتی	Injury Severity Score
کیفیت زیست	Sickness Impact Profile
Ulcer bleeding	Gastrointestinal Quality of Life Index
Appdlicitis	Forrest Classification
Pattern of organ injury	امتیازبندی های منتشر شده اما کم شناخته شده
شکستگی هیپ	Alvarado Appendicitis Score
کیفیت زیست	Moore Trauma Score
	Garden Classification
	Cologne Quality of Life Inventory

تشریح و طراحی سیستم‌های امتیازبندی

باید برخی از معیارهای ساده برای سیستم درجه‌بندی و مقیاس‌بندی به اجرا درآیند:

- ۱- سیستم باید نوع واحدی از پدیده‌های کیفی را عنوان کند، به این ترتیب که متغیرهای یکسانی برای کمیت‌ها و کیفیت‌ها داشته باشد. سبب و پرتغال را مانند چیزهای مشابه نمی‌توان با یک مقیاس شمرد.
- ۲- مقیاس باید به روشنی تعریف شده و در نظم سلسله مراتب رده‌بندی شود. Stage III می‌بایست براساس معیارهای بالینی یا ریاضی که به خوبی شناخته شده و مورد توافق هستند، بدتر یا شدیدتر از Stage I باشد.
- ۳- مرحله یا رده‌های مختلف باید به طور متقابل انحصاری باشند. یک بیمار با بیان مشخصی از یک بیماری باید براساس مجموعه‌ای از معیارها مثلاً به عنوان Stage II مرحله‌بندی شود، تداخل غیرقابل کنترل و تعریف نشده از مرحله می‌تواند سیستم طبقه‌بندی را مخدوش کند.
- ۴- یک سیستم باید جامع و دربرگیرنده همه انواع شدت یک بیماری از کمترین خطر تا بدترین مرحله از جمله همه انواع پدیده‌های جانبی باشد. مقیاس باید با حیطه‌سنجشی که قرار است بکار گیرد، منطبق باشد. برای بررسی عملکرد زانو در مردی ۳۵ ساله با فعالیت‌های تفریحی / ورزشی متعادل هر دو طبقه‌بندی Katz ADL و شاخص Tegner برای سنجش فعالیت‌های ورزشی نامناسبند، بیمار برای اندازه‌گیری با شاخص ADL، بسیار جوان و پرتحرک بوده و برای محاسبه به وسیله شاخص Tenger، فعالیت‌های ورزشی بیمار کافی نخواهد بود.
- ۵- پیچیدگی و سنگینی نسبی عناصر مختلف امتیاز باید به حساب آیند. یک امتیازبندی ساده مانند Breslow برای ملانوم بدخیم تنها ضخامت ملانوم را به میلی‌متر بررسی می‌کند. امتیازبندی پیچیده‌تری مانند طبقه‌بندی Child جنبه‌ها و توصیف‌کننده‌های موقعیت بیماری مانند: بیلی‌روبین، آلبومین، آسیت یا علائم عصبی و وضعیت تغذیه را به حساب می‌آورد. در میان پزشکان بالینی با تجربه، تمام این عوامل به طور یکسان، مهم و نمایانگر عملکرد کبد هستند. به طور مثال متغیرهای مقیاس کوما گلاسکو حرکات چشم‌ها، فعالیت اندام‌ها و واکنش کلامی را دربرمی‌گیرد. در هر حال پدیده‌ها در یک مثلاً اندازه‌گیری PH یا شاخص بررسی کیفیت زیست پیچیدگی بیشتری پیدا می‌کنند. کدام یک از این عوامل بعداً اهمیت بیشتری می‌یابد. علائم بیماری، موقعیت عاطفی یا عملکرد جسمانی؟ چه چیزی باید اندازه‌گیری شود یک مورد شایع یا متغیر یا موردی بسیار نادر اما دارای اهمیت؟ در مورد تخریب مخاط مری کدام یک مهم‌تر است: یک دوره طولانی رفلکس یا دوره‌های کوتاه متعدد با PH کمتر از ۴. این مطلب مشکل مربوط به قیاس کردن را به بحث در مورد طراحی امتیاز، می‌کشاند. ایجاد مقیاس‌های پیچیده و تجمع آنها که پدیده‌هایی مانند کیفیت زیست، ناتوانی با احساس رضایت از زندگی را ارائه می‌کند، موجب می‌شود که شخص با موضوعات مربوط به ساختار درونی یک امتیاز مواجه شود. به عنوان یادآوری باید گفت که براساس زمینه‌های آماری بیان شده با قابلیت‌های مشخص یا ضرورت‌های بالینی، تجربه و اعتبار سنگینی هر یک از وجوه یک مسئله برای امتیاز و درجه‌بندی توزین می‌شود، همچنان که این امر توسط شاخص Mannheim - Peritonitis انجام می‌پذیرد.

هدف یک امتیازبندی: توصیف، پیش‌بینی و ارزیابی

سه مشخصه اساسی یا هدف طراحی در یک سیستم امتیازبندی وجود دارد: توصیف، پیش‌بینی و ارزیابی. بسیاری از سیستم‌های امتیازبندی برای توصیف تشکیل شده‌اند که هدف آنها تقسیم‌بندی افراد یا پدیده‌ها به رده‌های مختلف براساس معیارهای مشخص در مقطعی از زمان است. این فرد با بیماری معینی در «کلاس III B» قرار می‌گیرد. مقیاس karnofsky برای بیماران سرطانی و مقیاس صدمات بافتی به وسیله Moore از این نوع سیستم‌های طبقه‌بندی آماری هستند.

امتیازبندی براساس:	منجر به:
توصیف	آزمون تشخیص
ارزیابی	درمان
پیش‌بینی	پیش‌بینی

هدف بعدی اتصال نتایج با سیستم طبقه‌بندی است. این نتایج می‌توانند تشخیصی، درمانی یا پیش‌آگهی‌دهنده باشند. به عنوان مثال بیمار مبتلا به **Forrest Ib ulcer Bleeding** نیاز به آندوسکوپی و اندازه‌گیری **Central Venous Pressure**، به عنوان اقدامات تشخیصی در مراحل بعدی و در صورتی که سالخورده و آسیب‌پذیر باشد، نیاز به عملیات انتخابی زود هنگام به عنوان اقدامات درمانی دارد که به وسیله پیش‌آگهی خوب یا بد به دنبال می‌شود. یک سیستم مطلوب امتیازبندی با تدابیر درمانی تشخیصی و پیش‌آگهی نتایج، در ارتباط است. هر قدر این ارتباط و اتصال بهتر باشد، موارد استفاده سیستم بیشتر خواهد بود. مقیاس‌های بدون ارتباط یا دارای ارتباط ضعیف با نتایج و پیامدها، «اشعار علمی» هستند. امتیازهای آپاندیسیت چنین نمونه‌هایی هستند.

سیستم‌های امتیازبندی پیش‌بینی‌کننده و توصیفی هم چنین ابزارهای اجتناب‌ناپذیر طبقه‌بندی بیماران در کارآزمایی‌های کنترل شده بالینی می‌باشند. بدون طبقه‌بندی مناسب براساس سیستم مرحله‌بندی، کلیه بیمارانی که در موقعیت «بد» هستند باید در یک شاخه از مطالعه گنجانده شوند، در حالی که بیماران دارای موقعیت «خوب» در شاخه دوم جای می‌گیرند.

سومین و مهمترین هدف سیستم امتیازبندی یا مرحله‌بندی، ارزیابی تغییرات در طول زمان است. سیستم ارزیابی‌کننده باید قادر به سنجش تغییر یا ثبات به روشی بسیار حساس در طول زمان باشد. در بسیاری از تحقیقات و موقعیت‌های مربوط به طبابت روزانه، پزشکان نسبت به تغییرات ایجاد شده پس از درمان، علاقه نشان می‌دهند. سیستم امتیازبندی مفید باید در برابر شناسایی این تغییرات پاسخگو باشد. نفوذ یک روش درمانی بر روی سیستم امتیازبندی باید به اندازه کافی وسیع بوده و طراحی امتیازها باید برای بیان تغییرات به صورت ارقام، به اندازه کافی حساس باشد. بسیاری از مثال‌های خوب بر این مطلب تأکید می‌ورزند. «شرح حال تأثیر بیماری» نگارش ماریلین برگر (Marylin Berner) فهرست بلندی از سئوالات است که بیماری یک بیمار را به روشی بسیار پیچیده و جامع توصیف می‌کند. اگر یک مداخله درمانی، بیماری یک بیمار را کاهش دهد حتی اگر شرایط بیمار را واقعاً بهبود بخشد، تنها تعداد معدودی از موارد تغییر می‌کند و حداقل تأثیر عددی بر روی مجموعه امتیازها پدید می‌آید، در این صورت هیچ یک از ابعاد تأثیر بیماری بر روی زندگی از بین نخواهد رفت. به همین جهت مطالعات منتشر شده این سؤال را مطرح می‌کنند که آیا فهرست «شرح حال تأثیر بیماری» قادر به سنجش کلیه تغییرات می‌باشد یا خیر؟ تأثیر مشابهی در «فهرست کیفیت زیست کولوگن» (Cologne Quality of Life Inventory) که فهرستی است شامل بیش از ۲۰۰ سؤال در ارتباط با کیفیت زیست، دیده می‌شود. ابزاری نیز به وسیله متخصصین خون و شیمی‌درمانی در دانشگاه کلن (Cologne) تهیه شده است که مجدداً این سؤال را مطرح می‌کند که آیا در صورت ایجاد تغییر در موارد معدود، بهبود کیفیت زیست یک بیمار به وسیله مجموعه امتیازهای یک ابزار سنجش بازنتاب خواهد یافت یا خیر؟ از سوی دیگر ابزارهایی برای سنجش وجود دارند که برای بیان تغییرات شرایط بیمار زیاده از حد بزرگند. شاخص کیفیت زیست اسپیتز (Spitzer Quality of Life Index) برای سنجش تغییرات بالینی در موقعیت بیمار، پس از انجام جراحی سرطان معده یا مری در طول دوره بعد از عمل، بسیار کلی است. تنها زمانی که بیمار بر اثر پیشرفت بیماری رو به مرگ است، تغییرات عددی شاخص اسپیتز قابل مشاهده است. به طور خلاصه تغییرات بالینی یا حیطة مورد نظر باید شناسایی شود و سیستم امتیازبندی باید برای سنجش تغییرات در این حیطة به شکل مناسبی طراحی شود.

معیار طراحی

برخی معیارهای طراحی برای انجام آزمایش‌های تشخیصی و سیستم‌های امتیازبندی باید به ایجاز ذکر شوند.

معیار یک سیستم امتیازبندی
قابلیت تولید
اعتبار
حساسیت

قابلیت تولید به این معنا است که ابزار امتیاز تنها حداقل تغییرات منطقی را در صورت بکارگیری مکرر این ابزار در بیمار با ثبات از نظر بالینی، نشان می‌دهد. بدون قابلیت تولید مناسب، (از لحاظ بیان آماری به عنوان تفاوت بین اندازه‌گیری‌ها و توزیع آنها،

محاسبه آماري (Kappa) آزمون تشخیصی تنها یک لاتاری است. جنبه مشاهده‌کننده درونی و سایر منابع تورش باید مورد توجه قرار گیرند.

اعتبار

معیار مهم بعدی نشان می‌دهد که آیا آزمون آن چه را که باید مورد سنجش قرار دهد، واقعاً اندازه‌گیری می‌کند یا خیر؟ اعتبار یک آزمون جدید هم می‌تواند از طریق مقایسه با یک استاندارد طلایی پذیرفته شده، مورد بررسی قرار گیرد (به عبارت دیگر مقایسه یک آزمون آزمایشگاهی جدید با یک آزمون دقیق و به ثبات رسیده اما پرهزینه) یا از طریق برقراری ارتباط منطقی با معیاری شناخته شده و به اثبات رسیده، به عنوان مثال بیماری دارای کیفیت زیست بسیار بهتر باید به طور موازی از درجه بالای تحرک جسمانی در منزل، در حوالی منزل و در سطح شهر نیز برخوردار باشد. در ساختار دورانی یک معیار با معیار دیگر که متغیری موازی و قابل قیاس است، در ارتباط بوده و مقایسه می‌شود. واژه‌هایی همچون «اعتبار ساختاری» و «اعتبار ظاهری» بیانگر این ارتباط است، در این زمینه باید متون نگارش یافته کارشناسان روش‌شناسی مورد مطالعه قرار گیرد.

سومین معیار طراحی حساسیت است که به معنای توانایی شاخص در انتخاب تغییرات بالینی می‌باشد. این مبحث در بخش مربوط به هدف ارزیابی امتیاز، مورد بررسی قرار گرفته است.

امتیازهای انتقالی یا امتیازهای ویژه بیماران

در امتیازهای انتقالی بیمار باید این مطلب را ثبت کند که موقعیتش نسبت به اندازه‌گیری قبلی بهتر یا بدتر شده است. این رویکرد موجب اجتناب از تعریفی روشن از آنچه که اکنون باید اندازه‌گیری شود، می‌گردد. بیمار به سادگی می‌گوید بهتر یا بدتر است. مشکل زمانی بروز می‌کند که اندازه‌گیری‌ها تکرار شده و مورد تفسیر قرار می‌گیرند و مقایسه بین افراد یا گروه‌هایی از افراد برقرار می‌شود. این سؤال مطرح می‌شود که آیا اطلاعات بدتر یا بهتر بین بیماران مختلف قابل مقایسه است یا خیر؟ و یا تفاوت‌های بین افراد فاحش است یا خیر؟ روش دیگر تعدیل امتیاز از طریق تخصیص یا انطباق بیماران صورت می‌گیرد به عنوان مثال برای یک بیمار، مطالعه، استراحت، خواب و اجتناب از حرکات‌های جسمانی می‌تواند جنبه بسیار مهمی از کیفیت زیست او باشد، در صورتی که برای بیمار دیگر فعالیت‌های ورزشی از جمله شرکت در مسابقات در کیفیت زیست او از اهمیت فوق‌العاده برخوردار است. سنجش کیفیت زیست به شکلی منطقی و با شاخصی مشابه در این دو فرد کار دشواری است. این مشکل را می‌توان از طریق انطباق یک سیستم امتیازبندی بر نیازهای بیمار به منظور سنجش و بیان این که آیا مداخله دارویی کیفیت زیست را بهبود بخشیده است یا خیر؟ حل کرد. «آیا ترمیم رباط متقاطع در زانو فعالیت ورزشی شما را بهبود بخشیده است؟ بسیار زیاد، زیاد، اندکی، به هیچ وجه بدتر شده است.» این مثال هم چنین نشان می‌دهد که طیف پاسخ‌های ارائه شده بر نتیجه حاصل از امتیازبندی تأثیر می‌گذارد و تفاوت فاحشی بین پاسخ‌های ارائه شده از «بسیار زیاد» تا «به هیچ وجه» و این که آیا جنبه منفی را نیز از طریق ارائه پاسخ «بسیار زیاد» تا «به شکل قابل ملاحظه‌ای بدتر» نیز در نظر می‌گیرد، ایجاد می‌کند. اندازه ماهی را که صید می‌کنید، مسلماً بستگی به اندازه سوراخ تور ماهیگیری شما دارد (این مطلب در مورد پاسخ‌ها هم صدق می‌کند). مجدداً یادآوری می‌شود با این امتیازهای منطبق با بیماران می‌توانید به درون مشکل مقایسه بین افراد و گروه‌هایی از افراد رخنه کنید.

محقق بالینی باید به خوبی مطلع باشد که روش، سیستم مرحله‌بندی، امتیاز و آزمون آزمایشگاهی، همگی بر نتیجه تأثیر می‌گذارند. تنها چیزی را بدست می‌آورید که در پی آن هستید. بنابراین آزمون و تخصیص آن باید قبل از شروع اندازه‌گیری به خوبی انتخاب شود. جای تعجب نیست که طراحی و انجام کارآزمایی تصادفی می‌تواند سال‌ها بطول انجامد.

در این جا سؤال‌های مفیدی که هنگام مواجهه با انتخاب سیستم امتیازبندی مناسب از خود می‌پرسید، بیان می‌شود. برخی سیستم‌های طبقه‌بندی مانند Child یا Dukes به حدی شناخته شده‌اند که بکارگیری آنها توصیه می‌شود. برای امتیازبندی‌هایی که کمتر شناخته شده هستند، سؤال‌های زیر باید با مراجعه به منابع اصلی و بکارگیری بعدی امتیازها پاسخ داده شوند. هر قدر به ابزاری مناسب با اهداف مورد نظر مطالعه نزدیک باشیم، می‌توانیم به بیشتر این سؤال‌ها پاسخ مثبت ارائه کنیم.

۱- آیا هدف امتیاز از سوی نگارنده به روشنی تعریف شده است؟

۲- امتیاز چگونه طراحی و ایجاد شده است؟

- ۳- آیا امتیاز از ساختار بالینی براساس تجربه، اعتبار، دانش و ضرورت‌ها برخوردار است و یا دارای ساختار ریاضی با کلیه جزئیات آماری و نیازها است؟
- ۴- نگارنده چه اندازه زمان، جزئیات و فعالیت‌های تحقیقاتی را برای ایجاد سیستم امتیازبندی صرف کرده است؟
- ۵- آیا اعتبار امتیاز بعدی از طریق یک روند رسمی و به اثبات رسیده اعتبار، مقایسه با سایر مقیاس‌ها و بکارگیری در سایر مطالب منتشر شده، تأیید شده است؟
- ۶- آیا امتیازبندی به وسیله محقق دیگری و در درون جمعیت مورد مطالعه دیگری غیر از جمعیتی که برای آن ایجاد گردیده، تعیین اعتبار شده است؟
- ۷- آیا امتیازبندی به شکلی موفقیت‌آمیز به وسیله محقق دیگری بکار گرفته شده است؟ و در سطحی وسیع در متون پزشکی منتشر شده است؟
- ۸- آیا امتیازبندی تنها توصیفی است و یا پیامدهایی برای رژیم غذایی، تشخیصی و درمانی بیمار که منجر به نتایج و پیش‌آگهی‌های متفاوت می‌شود داراست؟
- ۹- آیا امتیازبندی دارای طبقه‌بندی و طراحی مناسب برای سؤال‌های تحقیقاتی مطرح شده است؟
- ۱۰- آیا محاسبه امتیازبندی روشن است؟ کیفیت عددی آن چگونه است؟ آیا امتیاز برای محاسبه کامپیوتری مهیا شده و هدایت آن ساده است و می‌تواند به راحتی مورد استفاده قرار گیرد؟ تجزیه و تحلیل آماری چگونه است؟ نتایج چگونه تفسیر می‌شوند؟
- گاهی به طراحی یک سیستم امتیازبندی جدید بروز می‌کند، زیرا نمی‌توانید در متون مربوطه، امتیازبندی منطبق بر اهداف خود را بیابید. طراحی امتیازبندی جدید و شخصی، کار ساده‌ای نیست و به زمان، تجربه و پشتیبانی کارشناسان روش‌شناسی دارد. به هر حال همان‌گونه که فن‌شتاین (Feinstein) اشاره کرده، پزشکان بالینی ضمن این که باید خود تحقیقات پایه‌ایشان را انجام دهند، به پشتیبانی دانشمندان نظریه‌پرداز نیز نیازمندند. طراحی یک ابزار سنجش کار ساده‌ای نیست که در بعد از ظهر یک یکشنبه بتوان آن را انجام داد. این روند در توصیف شده است. انجام این کار نیاز به صرف وقت دارد، اما زمانی که قرار است یک سیستم امتیازبندی خوب ایجاد شود ارزش تلاش را دارد. اشارات و سؤال‌های مفید زیر انجام این روند را تسهیل می‌کند.
- ۱- آیا واقعاً سیستم امتیازبندی جدیدی مورد نیاز است یا تغییری کوچک و اعتبار بخشیدن مجدد به ابزار موجود، انتخابی درست است؟
- ۲- هدف مورد نظر در سیستم امتیازبندی جدید چیست؟
- ۳- جمعیت ویژه بیماران تحت بررسی کدام است؟ کدام گروه از بیماران مورد بررسی قرار خواهد گرفت؟
- ۴- چه میزان و چه نوعی از کمک روش‌شناختی مورد نیاز است؟
- بررسی دقیق این سؤال‌ها و پیروی از روند رسمی ارائه شده، کمک‌کننده است.

مثال‌هایی از سیستم امتیازبندی در موقعیت‌های بالینی

توصیفی کوتاه از سیستم‌های امتیازبندی و مرحله‌بندی، بر اهمیت این سیستم‌ها در کمک به حل مشکلات بالینی و پاسخگوئی به سؤال‌های مهم، تأکید می‌ورزد.

سرطان معده: در جراحی رایج سرطان معده کاهش در تشریح تمديد غدد لنفاوی ایجاد شده است. کارآزمایی‌های تصادفی اخیر در زمینه Radical Lymphadenectomy، افزایش میزان مرگ و میر و ناتوانی، بدون شانس بقا را نشان می‌دهد. در طول دو دهه اخیر هیچ‌گونه ابزار پذیرفته شده‌ای برای بررسی کیفیت زیست قبل و بعد از انجام عمل جراحی سرطان معده، وجود نداشت. به منظور انتخاب بیمار برای روش‌های درمانی چند وجهی یا خودداری از درمان با در نظر گرفتن میزان پیشرفت بیماری و راه عبور غذا پس از برداشتن معده، یک سیستم مفید امتیازبندی کیفیت زیست در سطح وسیعی حائز اهمیت بود. از طریق امتیازبندی ترویدل کوش (Trodel-Kusch) و در سال ۱۹۸۷ و همچنین شاخص کیفیت زیست گوارشی نشان داده شد که بیماران با Pouch Reconstruction بعد از انجام برداشتن معده دارای بهبود کیفیت زیست در سال‌های اولیه پس از عمل جراحی هستند. دوره‌ای دارای اهمیت فوق‌العاده زیرا ۴۰ تا ۶۰ درصد بیماران در طول این سال‌ها جان خود را از دست می‌دهند.

جراحی زانو: یکی از رایج‌ترین آسیب‌دیدگی‌های زانو، پارگی رباط‌های متقاطع داخلی است. این آسیب‌دیدگی نیاز به عمل جراحی فنی و دشواری دارد که امروزه از طریق آرتروسکوپی انجام می‌شود. چندین ماه طول می‌کشد تا عملکرد زانو به شکلی مناسب و بدون درد، مجدداً بدست آید. هم ارزیابی ذهنی (بررسی شخصی بیمار) و هم ارزیابی عینی (بررسی مصاحبه‌کننده + / - آزمون‌های بالینی) در این زمینه بکار گرفته شده‌اند. ابزار عملی بررسی عملکرد زانو براساس ارزیابی ذهنی، امتیازبندی فلاندری (Flandry) است، که اعتبار لازم، به ویژه در آلمان به آن داده نشده است. گروهی در بیمارستان کلن به ترجمه و ارزیابی مجدد این امتیازبندی پرداخته‌اند. با استفاده از این شاخص پس از ترمیم و بازتوانی رباط متقاطع قدامی، روند بالینی می‌تواند از طریق یک نمودار قابل درک گرافیکی که به راحتی برای بیمار و پزشک قابل توضیح است، به تصویر کشیده شود. در آینده این ابزار امکان مقایسه بین روش‌های درمانی مختلف (آرتروتومی در برابر آرتروسکوپی) را برای بیمار پدید آورد.

پیوند کبد: پیوند کبد یکی از پرهزینه‌ترین مداخلات درمانی در جراحی عمومی به شمار می‌رود. این‌گونه عمل‌های جراحی تنها در صورتی که منفعی مانند بقاء و کیفیت زیست را تضمین کنند، می‌توانند توجیه شوند.

کیفیت زیست در بیماری‌های کبدی به خوبی مورد بررسی قرار نگرفته است، این امر تا حدی بستگی به کاهش توانایی‌های روانی در بیماران مبتلا به نارسایی پیشرفته کبد، دارد. استفاده از «شاخص کیفیت زیست کبد» نشان داد که بیمارانی که تحت پیوند کبد Orthotopic قرار گرفته‌اند نسبت به بیمارانی که این عمل جراحی بر روی آنها انجام نشده است، از کیفیت زیست به مراتب بهتری برخوردارند. چنین اطلاعاتی برای ارتقاء برنامه پیوند کبد ضروری است.

عبارت زیر که از مک داوول (McDowell) و نول (Newell) نقل می‌شود، اهمیت و دشواری انتخاب یک ابزار سنجش مناسب را به طور خلاصه بیان می‌کند:

سرانجام، انتخاب یک ابزار سنجش عنصری از هنر و یا حتی نیک‌بختی در خود دارد، غالباً بکارگیری بیش از یک ابزار سنجش در صورت امکان، نهایت دوراندیشی است و این مزیت را دارد که زمانی که نتایج حاصله از دو روش به ظاهر مشابه مورد توافق باشند، نتیجه‌گیری مطالعه را تقویت می‌کنند. در عین حال درک کلی ما را از قابلیت مقایسه ابزار سنجشی که مورد استفاده ما است، افزایش می‌دهد.

تفسیر

معیار دیگر ارزیابی قابلیت استفاده سیستم امتیازبندی، حساسیت است «تلفیقی از شعور و دانشی قابل قبول در زمینه پاتوفیزیولوژی و واقعیت‌های بالینی» به عنوان مثال شاخصی برای سنجش بیماری التهابی روده که سؤال‌هایی در مورد سر درد را در برمی‌گیرد و سؤال‌های مربوط به روده را حذف می‌کند، غالباً فاقد حساسیت است.