

بررسی اعتبار اندازه‌گیری دندانپزشکان در ارزیابی شاخص نیاز به درمان ارتودنسی با استفاده از مدل‌بندی توافق

دکتر علیرضا کبیرزاده باغبان^{*}، دکتر غلامرضا بابایی، دکتر انوشیروان کاظم‌نژاد^{**}، دکتر سقراط فقیه‌زاده^{**}، دکتر سیدمحمد رضا صفوی^{***}، دکتر مهتاب نوری^{***}، دکتر علی‌اکبر سفیدرودی^{****}

چکیده

زمینه و هدف: معمولاً دندانپزشکان در تشخیص نیاز به درمان ارتودنسی برای گروه مشترکی از افراد، دارای نظرات متفاوتی نسبت به استاندارد طلایی هستند. برای بررسی میزان تطابق اندازه‌های حاصل توسط دندانپزشکان با استاندارد طلایی و تعیین اعتبار اینگونه اندازه‌گیریها که دارای مقیاس ترتیبی می‌باشند، در آمار روشهایی از قبیل محاسبه ضرایب توافق و همچنین مدل‌بندی توافق موجود است. در این مقاله ابتدا مشکلات ضرایب توافق از قبیل کاپا و کاپا وزندار بیان گردیده و سپس بررسی اعتبار داده‌ها از طریق مدل‌بندی انجام می‌شود. هدف از انجام این کار ارائه جایگزین مناسبتر برای ضرایب توافق می‌باشد.

روش بررسی: تحقیق حاضر که از نوع مقطعی می‌باشد بر روی داده‌های شاخص نیاز به درمان ارتودنسی انجام شده است. این شاخص که توسط یک دندانپزشک مجرب به عنوان استاندارد طلایی و ده دندانپزشک دیگر بطور مستقل، روی ۵۱ کست استاندارد مطالعه ارتودنسی با استفاده از خط کش استاندارد DHC اندازه‌گیری شده است دارای پنج سطح عدم نیاز به درمان، نیاز جزئی به درمان، نیاز متوسط به درمان، نیاز شدید به درمان و نیاز بسیار شدید به درمان می‌باشد، اما در این مقاله از شاخص اصلاح شده سه‌سطحی استفاده شده است. مدلهای آماری پارامتر قطری و پیوند یکنواخت نیز به عنوان مدلهای آماری مورد نظر مبنای استنباطها قرار گرفتند. در نهایت مدلهای مذکور به ۱۰ جدول حاصل از نرخ‌گذاری متقاطع هر کدام از ۱۰ دندانپزشک در مقایسه با استاندارد طلایی برازش داده شدند.

یافته‌ها: در ۶ مورد مدل پارامتر قطری (با حداقل سطح معنی‌داری ۰/۳۴) و در ۴ مورد دیگر مدل پیوند یکنواخت (با حداقل سطح معنی‌داری ۰/۳۶) برازش داشتند. نزدیکترین ارزیابی‌ها به استاندارد طلایی مربوط به دندانپزشکانی بود که برای آنها مدل آماری پیوند یکنواخت مناسب بود در حالی که دندانپزشکانی که برای آنها مدل آماری پارامتر قطری مناسب بود، نمونه‌ها را بسیار بالاتر از استاندارد طلایی ارزیابی کرده بودند.

نتیجه‌گیری: همه دندانپزشکان کستها را بالاتر از استاندارد طلایی ارزیابی کردند بطوریکه کمترین مقدار نسبت بخت موضعی برابر ۶/۴ و بیشترین مقدار آن برابر ۱۲۱/۵ بود، بنابراین آموزش همه آنها برای اندازه‌گیری شاخص نیاز به درمان ارتودنسی لازم به نظر می‌رسد.

کلید واژه‌ها: مدل‌بندی توافق، داده‌های ترتیبی، مدل آماری پیوند، شاخص نیاز به درمان ارتودنسی (IOTN)

تاریخ دریافت مقاله: ۱۳۸۳/۳/۲۶ تاریخ اصلاح نهایی: ۱۳۸۳/۶/۸ تاریخ تأیید مقاله: ۱۳۸۳/۷/۴

مقدمه

توافقی (contingency table) گنجانده در چنین گروه‌بندی‌هایی به ندرت توافق کاملی بین آزمایش‌ها وجود دارد، یعنی داده‌ها صددرصد قابل اعتماد نیستند. ارزیابی قابلیت اعتماد

فرض کنیم دو آزماینده بطور جداگانه هر کدام از افراد یک نمونه را طبق یک مقیاس ترتیبی (ordinal scale) طبقه‌بندی می‌کنند. می‌توان نرخ‌گذاری (rating) توأم آنها را در یک جدول

(Reliability)، همچنین ارزیابی اعتبار (Validity) از منظر توافق بین آزمایش‌ها با یک استاندارد طلایی (golden standard) قابل بررسی است. (۲،۱) ارزیابی اعتبار اندازه‌گیریها به این دلیل مهم است که حتی در صورت وجود توافق کامل میان آزمایش‌ها، ممکن است این داده‌ها با داده‌های حاصل از اندازه‌گیری استاندارد طلایی متفاوت باشند. بدین معنی که تمامی آزمایش‌ها در ارزیابی خود عمداً یا سهواً دارای جهت‌گیری مشخصی باشند.

یکی از موضوعاتی که می‌توان در چنین جدولی بررسی نمود، توافق بین آزمایش‌هاست. این موضوع در واقع فراوانی روی قطر اصلی در توزیع توام نرخ‌گذاریها می‌باشد. در این مقاله ابتدا مشکلات ضرایب توافق (coefficients of agreement) بیان گردیده، سپس با استفاده از مدل‌بندی ساختار توافق داده‌های حاصل از جدول‌بندی متقاطع تک‌تک آزمایش‌ها با استاندارد طلایی، توافق اندازه‌گیریها مورد بررسی قرار می‌گیرد. از آنجائیکه یکی از آزمایش‌ها استاندارد طلایی می‌باشد، این تحلیل را می‌توان معادل ارزیابی اعتبار داده‌ها در نظر گرفت. (۳) زمانی که متغیر کیفی باشد ساده‌ترین و متداول‌ترین راه برای بررسی توافق بین چند آزمایش یا بررسی توافق میان آزمایش‌ها با یک استاندارد طلایی، محاسبه شاخص‌های توافق بین نرخ‌گذاریهای آنها می‌باشد. در این میان ضرایب کاپا و کاپا وزن‌دار بیشتر از سایر شاخص‌ها مورد استفاده قرار می‌گیرند. (۵،۴) مقالات متعددی در خصوص مشکلات استفاده از این شاخص‌ها چاپ گردیده، به نقاط ضعف آنها پرداخته شده است. (۶) پاره ای از این مشکلات عبارتند از: ۱- این شاخص‌ها برای مقایسه فراوانی‌های مشاهده شده و مورد انتظار، مدل استقلال بین نرخ‌گذاریها را به عنوان رابطه مبنا مورد استفاده قرار می‌دهند. لذا در صورتیکه این فرض برقرار نباشد، استفاده از آنها می‌تواند گمراه‌کننده باشد و اتفاقاً در داده‌های با مقیاس ترتیبی به ندرت فرض استقلال برقرار است. (۱) ۲- ضریب

کاپا فقط سطح توافق روی هم مشاهده شده را با سطح توافق روی هم مورد انتظار مقایسه می‌کند و به الگوی توافق (نحوه پراکنندگی داده‌ها روی قطر اصلی جدول) کاری ندارد. (۷) ۳- این شاخص‌ها فقط برای جداول مربعی که در آنها هر دو آزمایش سطوح یکسانی از نرخ‌گذاری را تشخیص دهند قابل محاسبه می‌باشد. یعنی در صورتیکه یکی از سطوح مقیاس ترتیبی، توسط یک آزمایش تشخیص داده شود ولی توسط آزمایش دیگر تشخیص داده نشود، این ضرایب قابل محاسبه نخواهند بود. ۴- کاپا وزن‌دار که از آن فقط برای بررسی توافق بین نرخ‌گذاریهای ترتیبی استفاده می‌شود، بیشتر از نسبت دقیق توافق نسبت به انتخاب وزنها حساس است. (۱) به این معنی که با دادن مجموعه‌های متفاوتی از وزنها به سلول‌های جدول توافقی، مقادیر مختلفی برای این ضریب به دست می‌آیند، که در نتیجه تفسیر این ضریب دشوار می‌شود. ۵- این ضرایب تنها اندازه‌هایی از توافق بوده و هیچگونه اطلاعاتی در مورد ساختار یا الگوی توافق در اختیار نمی‌گذارند. (۱) بنابراین بهتر است به جای توصیف ساختار توافق داده‌های چنین جدولی توسط یک عدد، این ساختار را مدل‌بندی نماییم تا بتوانیم علاوه بر اندازه‌گیری میزان دوری یا نزدیکی اندازه‌های آزمایش‌ها نسبت به استاندارد طلایی، در مورد جهت نرخ‌گذاری آنها نسبت به استاندارد طلایی نیز اظهار نظر نماییم. بدین معنی که بررسی نماییم آزمایش‌ها نسبت به استاندارد طلایی، نمونه‌ها را بالاتر نرخ‌گذاری می‌کنند یا پایین‌تر. اولین بار این کار را Tanner و Young در ۱۹۸۵ انجام دادند. (۸) این بحث با برآورد معادله رگرسیون خطی برای دو متغیر کمی، بجای محاسبه ضریب همبستگی خطی پیرسن برای آن دو متغیر، قابل مقایسه است.

در این مقاله برای توصیف روش مدل‌بندی، ابتدا برای داده‌های مربوط به شاخص ترتیبی نیاز به درمان ارتودنسی، برازندگی (fitness) مدل‌های مورد نظر بررسی شده است. سپس بهترین

مدل انتخاب گردیده، پارامتر آن برآورد و تفسیر شده است. شاخص مذکور ۱۱ بار بطور مستقل (یک بار توسط یک دندانپزشک مجرب به عنوان استاندارد طلایی و ۱۰ بار توسط دندانپزشکان دیگر به عنوان آزماینده) روی ۵۱ کست استاندارد مطالعه ارتودنسی با استفاده از خطکش استاندارد DHC اندازه‌گیری شده است که نرخ‌گذاری متقاطع هر کدام از این آزماینده‌ها با استاندارد طلایی به ایجاد ۱۰ جدول منجر می‌شود. هدف این مطالعه آن است که با توجه به اشکالات وارده به ضرایب توافق، به کمک مدل‌های آماری مذکور که فاقد این مشکلات می‌باشند، در مورد وضعیت نرخ‌گذاری دندانپزشکان نسبت به استاندارد طلایی اظهار نظر نماییم. به عبارتی می‌خواهیم اعتبار اندازه‌گیری دندانپزشکان را با استفاده از این مدل‌ها ارزیابی نماییم.

روش بررسی

مطالعه انجام شده از نوع مقطعی تحلیلی (analytic cross-sectional) می‌باشد که در آن برای ارزیابی اعتبار اندازه‌گیری‌های شاخص نیاز به درمان ارتودنسی با استفاده از مدل‌بندی و رسیدن به جهت نرخ‌گذاری آزماینده‌ها نسبت به استاندارد طلایی، مدل‌های آماری پارامتر قطری (۸) (diagonal parameter model)، پیوند یکنواخت (۹) (uniform association model) و توافق بعلاوه پیوند یکنواخت (۱۱،۱۰) (agreement plus uniform association model) مورد نظر قرار گرفتند. لازم به ذکر است که همه مدل‌های مذکور جزو مدل‌های آماری لگ خطی (log-linear models) بوده و مدل‌های آماری پارامتر قطری و پیوند یکنواخت حالت خاصی از مدل توافق بعلاوه پیوند یکنواخت می‌باشند. (۱۱)

داده‌های این مقاله قسمتی از داده‌های مربوط به طرح "ارزیابی روایی و پایایی شاخص نیاز به درمان ارتودنسی، IOTN (Index

(of Orthodontic Treatment Need) است که در تابستان ۱۳۸۲ در مرکز تحقیقات علوم دندانپزشکی انجام شد. در یک مرحله از این طرح، هر کدام از ۱۰ دندانپزشک که به عنوان آزماینده حضور داشتند پس از طی کارگاه آموزشی ۴ روزه، این شاخص را بطور مستقل روی ۵۱ کست استاندارد مطالعه ارتودنسی - متعلق به بخش ارتودنسی دانشکده دندانپزشکی دانشگاه علوم پزشکی شهید بهشتی - با استفاده از خطکش استاندارد شفاف DHC - متعلق به مرکز تحقیقات علوم دندانپزشکی - اندازه‌گیری نمودند. داده‌ها در فرم اطلاعاتی مربوط ثبت گردیدند. بعلاوه جهت بررسی اعتبار داده‌های حاصل، یک دندانپزشک مجرب نیز به عنوان استاندارد طلایی، شاخص مذکور را روی این کست‌ها اندازه‌گیری نمود. هدف طرح این بود که در صورتی که اندازه‌گیری‌های دندانپزشکان معتبر باشد، این افراد شاخص مذکور را برای نوجوانان دبیرستانی ۲۰ منطقه شهر تهران برآورد نمایند. این شاخص که به کمک ۳۰ مشخصه و به صورت بیشترین امتیاز (maximum score) اندازه‌گیری می‌شود، رتبه‌ای بوده و دارای ۵ سطح عدم نیاز به درمان، نیاز جزئی به درمان، نیاز متوسط به درمان، نیاز شدید به درمان و نیاز بسیار شدید به درمان می‌باشد و در بسیاری از کشورهای اروپایی و آسیایی - که در آنها بیمه درمان ارتودنسی مطرح می‌باشد - مورد استفاده قرار می‌گیرد. لازم به ذکر است که این شاخص بطور عمده برای اندازه‌گیری نیاز به درمان ارتودنسی در سطح جامعه کاربرد دارد. (۱۲-۱۴) در این پژوهش از شاخص اصلاح شده نیاز به درمان ارتودنسی (۱۵) با سه سطح عدم نیاز یا نیاز جزئی به درمان (رتبه‌های ۱ و ۲)، نیاز متوسط به درمان (رتبه ۳) و نیاز قطعی به درمان (رتبه‌های ۴ و ۵) استفاده شد و در نهایت ۱۰ جدول ۳×۳ - که نتیجه جدول‌بندی متقاطع حاصل از نرخ‌گذاری هر یک از ۱۰ دندانپزشک به عنوان آزماینده در مقایسه با دندانپزشک مجرب به عنوان استاندارد طلایی می‌باشد - مبنای عمل قرار گرفتند.

یافته‌ها

در بررسی مناسبت مدل‌های مورد نظر روی مجموعه داده‌های مذکور، مدل توافق بعلاوه پیوند یکنواخت برای هیچکدام از ۱۰ جدول مناسب نبود. در ۴ مورد مدل پیوند یکنواخت و در ۶ مورد دیگر مدل پارامتر قطری بهترین مدلها تشخیص داده شدند. جدول ۱ خلاصه‌ای از کارهای انجام شده روی این مجموعه داده را نشان می‌دهد. لازم به ذکر است که تعدادی از سلولهای مربوط به هر کدام از این ۱۰ جدول خالی بود (دارای فراوانی مشاهده شده صفر بود). این وضعیت عموماً برای چنین مجموعه داده‌هایی اتفاق می‌افتد. زمانی که آزمونهای نیکویی برازش برای چنین داده‌هایی به کار برده می‌شوند، اگر چه توزیع تقریبی این آماره دقیقاً خی-دو (chi-square) نمی‌شود اما این آماره برای مقایسه مدلها خوب عمل می‌کند. (۱۱) بنابراین می‌توان از آن به عنوان ابزاری برای مقایسه مدلها و انتخاب مناسب‌ترین مدل استفاده نمود.

همانطور که قبلاً نیز عنوان شد ستونهای ۳ تا ۷ جدول به کمک نرم افزار SAS و ستون ۸ بصورت دستی و از روی ستون ۵ محاسبه شده‌اند. بر اساس یافته‌ها، همه مدلها دارای برازندگی مناسب می‌باشند زیرا همه مقادیر احتمال موجود در ستون ۴ جدول از ۰/۰۵ بزرگترند. قویترین مدل در بین ۱۰ مدل موجود دارای کمترین مقدار آماره نیکویی برازش ۱/۳۸ و بیشترین سطح معنی‌داری ۰/۷۱ می‌باشد در حالیکه برای ضعیف‌ترین مدل موجود مقدار آماره نیکویی برازش و سطح معنی‌داری به ترتیب عبارتند از: ۳/۳۵ و ۰/۳۴.

تعداد درجات آزادی برای بررسی مناسبت هر دو مدل مطرح شده در بالا در هر ۱۰ جدول ۳ می‌باشد. اگر پارامترهای معرف جمله (عبارت) توافق در مدلهای پیوند یکنواخت و پارامتر قطری را به ترتیب با β و δ نشان دهیم، ستون ۵ جدول فوق مقادیر برآورد شده این پارامترها (یعنی β و δ) و ستون ۷ معنی‌داری آنها را در بر دارد و از آنجائیکه همه این مقادیر

در این راستا برای هر کدام از ۱۰ جدول فوق مناسب (goodness) مدل آماری پارامتر قطری و بعد از آن مدل پیوند یکنواخت و در نهایت مدل توافق بعلاوه پیوند یکنواخت بررسی شد تا بهترین مدل آماری ممکن انتخاب گردد. ملاک مقایسه مدلها و انتخاب آنها، مقادیر آماره نیکویی برازش (goodness of fit statistics) و سطح معنی‌داری این آماره‌ها بود. بدین معنی که بعد از برازش مدلها، آنهایی مورد نظر قرار می‌گرفتند که سطح معنی‌داری آماره نیکویی برازش آنها حداقل ۰/۰۵ بود. بعلاوه در صورتی که برای بعضی از این جداول بیش از یک مدل مناسب بود، برای انتخاب بهترین مدل، آن مدلی مورد قبول قرار گرفت که مقدار آماره نیکویی برازش کمتر و در نتیجه سطح معنی‌داری بالاتری داشت. در گام بعد و پس از رسیدن به بهترین مدل، معنی‌داری عبارت (جمله) تعیین‌کننده توافق در مدل بررسی گردید. برای این منظور در صورتی این عبارت معنی‌دار در نظر گرفته شد که سطح معنی‌داری آن حد اکثر ۰/۰۵ باشد. در نهایت برای بررسی وضعیت نرخ گذاری هر کدام از دندانپزشکان نسبت به استاندارد طلایی از نسبت بخت موضعی (local odds ratio)، که از روی همین مدلها بدست می‌آید، استفاده شد. (۱۱) برای استخراج جداول دو بعدی از نرم افزار SPSS ۱۲ و برای برازش مدلهای مورد نظر روی داده‌های جداول مذکور و برآورد پارامترهای این مدلها - با روش حداکثر درستنمایی (maximum likelihood) - از نرم افزار SAS ۸ استفاده گردید. اگر چه ممکن است تعداد ۵۱ کست به عنوان نمونه مورد بررسی، ظاهراً تعداد کمی باشد، اما طبق منابع موجود چون حجم نمونه در مطالعات مربوط به بررسی روایی و پایایی کم است (۱۵ الی ۲۰ نمونه برای متغیرهای کمی و اندکی بیشتر برای متغیرهای کیفی) (۱۶)، لذا تعداد نمونه‌ها در این تحقیق مناسب بوده و می‌تواند زیربنای انجام استنباطهای آماری قرار گیرد.

احتمال از ۰/۰۵ کمتر می‌باشند، این پارامترها از نظر آماری معنی‌دار بوده و باید در مدلها باشند.

جدول ۱- بهترین مدل برازش شده و جزئیات مربوط به آن برای هر کدام از ۱۰

جدول ۳ × ۳ حاصل از نرخ گذاری دندانپزشکان با استاندارد طلایی

شماره ردیف	بهترین مدل برازش شده	مقدار آماره نیکویی	مقدار احتمال مربوط به	مقدار برآورد شده پارامتر	مقدار آماره مرتبط با	مقدار احتمال مرتبط با پارامتر	نسبت بخت موضعی
(ستون ۱)	(ستون ۲)	(ستون ۳)	(ستون ۴)	(ستون ۵)	(ستون ۶)	(ستون ۷)	(ستون ۸)
۱	پیوند یکنواخت	۱/۶۲	۰/۶۵	۲/۸۷	۵/۵۵	۰/۰۱۸۵	۱۷/۶
۲	"	۲/۴۹	۰/۴۸	۲/۱۸	۵/۹۳	۰/۰۱۴۹	۸/۸
۳	"	۱/۳۸	۰/۷۱	۴/۱۷	۱۷/۵۱	۰/۰۰۰۱	۶۴/۷
۴	"	۳/۲۰	۰/۳۶	۱/۸۵	۴/۸۱	۰/۰۲۸۴	۶/۴
۵	پارامتر قطری	۲/۱۱	۰/۵۵	۲/۴۰	۱۷/۹۰	۰/۰۰۰۱	۱۲۱/۵
۶	"	۲/۰۸	۰/۵۶	۲/۱۶	۱۸/۹۹	۰/۰۰۰۱	۷۵/۲
۷	"	۲/۸۸	۰/۴۱	۱/۸۷	۱۵/۶۰	۰/۰۰۰۱	۴۲/۱
۸	"	۲/۳۸	۰/۵۰	۲/۰۹	۱۸/۲۸	۰/۰۰۰۱	۶۵/۴
۹	"	۲/۱۲	۰/۵۵	۲/۳۹	۱۷/۷۰	۰/۰۰۰۱	۱۱۹/۱
۱۰	"	۳/۳۴	۰/۳۴	۱/۶۰	۱۳/۸۰	۰/۰۰۰۲	۲۴/۵

همه دندانپزشکان به نرخ گذاری بالاتر نسبت به استاندارد طلایی تمایل دارند و در هیچ موردی نرخ گذاری آنها پایین تر از استاندارد طلایی نمی‌باشد. به این ترتیب می‌توان گفت که دندانپزشک ۵ با مقدار برآورد شده ۱۲۱/۵ بیشترین تمایل به نرخ گذاری بالا و دندانپزشک ۴ با مقدار برآورد شده ۶/۴ کمترین تمایل به نرخ گذاری بالا را دارند و سایر دندانپزشکان بین این دامنه قرار دارند. برای این دو دندانپزشک به ترتیب مدل‌های پیوند یکنواخت و پارامتر قطری مناسب بودند.

اگر چه هر ۴ دندانپزشکی که برای آنها مدل پیوند یکنواخت بهترین مدل تشخیص داده شد، دارای نسبت بخت موضعی بالاتر از ۱ می‌باشند ولی در اغلب موارد این اعداد خیلی بزرگ نبوده و بیانگر نزدیکی اندازه‌های این دندانپزشکان به استاندارد طلایی می‌باشد. در ۶ مورد دیگر که مدل پارامتر قطری مناسب تشخیص داده شده است، نسبت بخت موضعی در اغلب

ستون ۸ جدول فوق برای مدل پیوند یکنواخت از طریق $\exp(\beta)$ و برای مدل پارامتر قطری از طریق $\exp(2\delta)$ محاسبه گردید که در آنها β و δ به ترتیب برآوردهای β و δ در مدل‌های پیوند یکنواخت و پارامتر قطری می‌باشند. (۱۱) این مقادیر معرف نسبت بخت گروه‌بندی سطح ۲ نسبت به سطح ۱ و همچنین سطح ۳ نسبت به سطح ۲ برای هر کدام از ۱۰ دندانپزشک نسبت به استاندارد طلایی می‌باشند.

بحث

برای بررسی جهت توافق نرخ گذاری‌های هر کدام از دندانپزشکان نسبت به استاندارد طلایی و ارزیابی اعتبار آنها باید به نسبت بخت موضعی مراجعه نمود. همانطور که دیده می‌شود نسبت بخت موضعی برای همه دندانپزشکان بطور قابل توجهی از عدد ۱ بالاتر است. این بدان معنی است که

استفاده نمود.

نتیجه گیری

نسبت بخت موضعی برای دندانپزشکان بین ۶/۴ و ۱۲۱/۵ در تغییر بود. لذا اگرچه عملکرد آنها در اندازه‌گیری شاخص نیاز به درمان ارتودنسی در مقایسه با هم متفاوت بود ولی همه آنها کسپتها را بالاتر از استاندارد طلایی نرخ‌گذاری نمودند. بنابراین برای اندازه‌گیری دقیق شاخص مذکور همه آنها به آموزش مجدد نیاز داشتند. یکی از دلایل این امر می‌تواند احتیاط دندانپزشکان باشد که منجر به انتساب سطوح بالای شاخص به نمونه‌ها می‌شود.

بنابراین در صورتی که بخواهیم توافق چند آزماینده با یک استاندارد طلایی را بررسی نموده و جهت نرخ‌گذاری آنها را مشخص نماییم، باید مدل‌های آماری پیوند یکنواخت، پارامتر قطری و توافق بعلاوه پیوند یکنواخت را برای مدل‌بندی ساختار توافق داده‌ها، مورد نظر قرار دهیم و بعد از برآورد پارامترهای این مدل‌ها به تفسیر ضرایب آنها بپردازیم تا آزماینده‌های مشکل‌دار و مشکلات نرخ‌گذاری آنها مشخص گردند. توصیه می‌شود در انجام این کار، مخصوصاً اگر تعداد سطوح متغیر زیاد باشد، تعداد نمونه در حد قابل قبولی باشند.

موارد بطور قابل ملاحظه‌ای از عدد ۱ بالاتر می‌باشد که این امر معرف دورتر بودن نرخ‌گذاری این دندانپزشکان نسبت به استاندارد طلایی است.

در بعضی منابع، مدل آماری توافق بعلاوه پیوند یکنواخت برای مدل‌بندی ساختار توافق داده‌های رتبه‌ای بهتر از مدل‌های پارامتر قطری و پیوند یکنواخت معرفی شده است. (۱۱،۱۰) اما با توجه به پیچیدگی آن به دلیل وجود پارامترهای بیشتر نسبت به دو مدل دیگر، این مدل به هیچ کدام از ۱۰ جدول مورد بررسی در این مقاله برازش پیدا نکرد.

در تحقیقات مشابه برای بررسی روایی و پایایی اندازه‌گیری نیاز به درمان ارتودنسی با استفاده از IOTN، از کاپا و کاپا وزندار استفاده شده است. (۱۷-۲۱) اگرچه این ضرایب میزان توافق را نشان می‌دهند اما با توجه به اشکالاتی که در مقالات مختلف به آنها وارد شده (۱-۷) و نظر به اینکه نمی‌توان به کمک آنها در مورد جهت ارزیابی آزماینده‌ها نسبت به استاندارد طلایی اظهار نظر نمود، استفاده از این ضرایب چندان خالی از ایراد نبوده و لازم است که با استفاده از مدل‌های آماری به این سوالات پاسخ گفت تا بتوان در مراحل بعدی آموزش از این یافته‌ها بهره برد. برای بررسی توافق نیز می‌توان از نسبت بخت موضعی حاصل از مدل‌های آماری، بجای ضرایب توافق

References

1. Perkins SM, Becker MP: Assessing rater agreement using marginal association models. *Stat Med* 2002; 21:1743-60.
2. Koch GG, Landis JR, Freeman JL, Freeman DH, Lehnen RG: A general methodology for the analysis of experiments with repeated measurement of categorical data. *Biometrics* 1977;33:133-58.
3. Kraemer HC, Periyakoil VS, Noda A: Tutorial in biostatistics, kappa coefficient in medical research. *Stat Med* 2002;21:2109-29.
4. Cohen J: A coefficient of agreement for nominal scales. *Education and Psychological Measures* 1960;20:37-46.
5. May SM: Modeling observer agreement – an alternative to kappa. *J Clin Epidemiol* 1994;44:1315-24.
6. Feinstein AR, Cicchetti DV: High agreement but low kappa: I. The problem of two paradoxes. *J Clin Epidemiol* 1990;43:543-549.

7. Light RJ: Measures of response agreement for qualitative data: some generalizations and alternatives. *Psychol Bull* 1971;5:365-77.
8. Tanner MA, Young MA: Modelling agreement among raters. *JASA* 1985;80:175-80.
9. Goodman LA: Simple models for the analysis of association in cross-classifications having ordered categories. *JASA* 1979;74:537-52.
10. Agresti A: *Categorical Data Analysis*. 2nd Ed. New York: John Wiley & Sons. 2002; Chap10:369.
11. Agresti A: A model for agreement between ratings on an ordinal scale. *Biometrics* 1988; 44:539-48.
12. Waring D, Jones JW: Does the GDP need to know about IOTN. *Dent Update* 2003;30:123-30.
13. Cooper S, Mandall NA, Dibiasi D, Shaw WC: The reliability of the Index of Orthodontic Treatment Need over time. *J Orthod* 2000;27:47-53.
14. Brook PH, Shaw WC: The development of an index for orthodontic treatment priority. *Eur J Orthod* 1989; 11:309-20.
15. Richmond S, Shaw WC, O'Brien KD et al: The relationship between the Index of Orthodontic Treatment Need and the consensus opinion of a panel of 74 dentists. *Br Dent J* 1995;178:370-4.
16. Fleiss JL: *The Design and Analysis of Clinical Experiments*. 1st Ed. New York: John Wiley & Sons. 1999; Chap1:8.
17. Jones CM, Wood K, O'Brien K, et al: Index of orthodontic treatment need, its use in a dental epidemiology survey calibration exercise. *Community Dent Health* 1996;13:208-10.
18. Richmond S, O'brien KD, Roberts CT, et al: Dentists variation in determination of Orthodontic Treatment Need. *Br J Orthod* 1994;21:65-8.
19. Richmond S, Roberts CT, Andrews M: Use of the Index of Orthodontic Treatment Need (IOTN) in assessing the need for Orthodontic Treatment pre- and - post - appliance therapy. *Br J Orthod* 1994; 21:175-84.
20. Burden DJ, Pine CM, Burnside G: Modified IOTN: an orthodontic treatment need index for use in oral health surveys. *Community Dent Oral Epidemiol* 2001;29: 220-5.
21. Kerosuo H, Enezi SA, Kerosuo E, Abdulkarim E: Association between normative and self perceived orthodontic treatment need among Arab high school students. *Am J Orthod Dentofacial Orthop* 2004;125:373-8.