

Standardy postępowania w badaniach naukowych

Część IV. Kryteria jakościowe oceny indeksów ubytków pochodzenia niepróchniczowego w aspekcie standaryzacji

Irena Zalewska¹ i Marta Tanasiewicz²

Standards of procedure in scientific studies. Part IV. Qualitative criteria in evaluation of indices for non-carious cavities from the aspect of standardization

Praca recenzowana

¹NZOZ Dental-Lux w Katowicach

Kierownik: lek. stom. Jolanta Suchacka-Niewęglowska

e-mail: irenazalewska-7@wp.pl

²Katedra Stomatologii Zachowawczej z Endodoncją Śląskiego Uniwersytetu Medycznego w Katowicach

Kierownik: dr hab. n. med. Marta Tanasiewicz

e-mail: martatanasiewicz@sum.edu.pl

Streszczenie

Głównym krokiem na drodze do stworzenia adekwatnego, uniwersalnego indeksu przeznaczanego do oceny ubytków niepróchniczych jest przeprowadzenie jego standaryzacji. W pracy położono nacisk na kryteria jakościowe, takie jak: rzetelność, trafność, czułość i swoistość, których zanalizowanie jest niezbędne do zrealizowania tego procesu. Opisano pojęcie zewnętrznego kryterium, tzw. złotego standardu, który stanowi punkt odniesienia dla wyników uzyskanych za pomocą analizowanego instrumentu badawczego. Przytoczono także dyskutowane w piśmiennictwie wymagania dotyczące indeksu „idealnego” oraz czynniki mogące utrudniać jego stworzenie.

Summary

The main step on the way to forming an adequate, universal index aimed at the evaluation of non-carious cavities is to carry out its standardization. In this study emphasis was placed on qualitative criteria such as reliability, accuracy, sensitivity and specificity, whose examination is necessary in order to achieve this process. A description is given of the concept of an external criterion, the so-called gold standard, that constitutes a point of reference for results obtained from the analysis of data obtained with the help of an analyzed study tool. Attention was also paid to requirements discussed in the literature that are needed for the „ideal” index together with the factors that may hinder its formulation.

Hasła indeksowe: ubytki niepróchnicze, rzetelność, trafność, czułość, swoistość

Key words: non carious defects, reliability, validity, sensitivity, specificity

W praktyce stomatologicznej coraz częściej obserwuje się u pacjentów ubytki pochodzenia niepróchniczowego. Istotne staje się zatem zagadnienie właściwej rejestracji danych dotyczących występowania tych defektów, stopnia ich nasilenia, prognozowania ewentualnego rozwoju, jak również zaplanowania właściwego leczenia. Ważną kwestią jest ponadto możliwość porównywania wyników uzyskanych z niezależnych badań (1-3).

W piśmiennictwie spotyka się wiele indeksów zarówno etiologicznych – jak indeksy erozji, a także tych pomijających czynnik sprawczy, np. indeksy starcia. Mogą one bazować na kryteriach ilościowych, do których zalicza się wysokość korony, głębokość czy rozległość ubytków, lub na podanych w formie opisowej kryteriach jakościowych. Stopień starcia jest wtedy określany mianem łagodnego, średniego lub ciężkiego (1, 4). Dotychczas jednak żaden z nich nie uzyskał miana złotego standardu w rejestracji ubytków niepróchniczych. Pojęcie złoty standard oznacza w tym wypadku

dostępny, uznany i szeroko akceptowany pomiar, metodę, narzędzie, lub procedurę (1, 3, 5). Istnieje potrzeba opracowania indeksu uniwersalnego, mającego zastosowanie zarówno w badaniach jednostki, jak i w badaniach epidemiologicznych (1, 3). Jest też konieczne uzyskanie odpowiedzi na pytania (1, 2, 4, 6, 7)

- Jakie cechy powinny charakteryzować indeks „idealny”?

- Jakie kryteria należy przyjąć, aby móc wiarygodnie i rzetelnie porównywać rezultaty różnych badań dotyczących tego samego zagadnienia?

- Jak uściślić standaryzację terminologii i definicje kryteriów oceny, aby uniknąć dowolności w interpretacji wyników?

Należy przy tym zaznaczyć, że w przypadku ubytków pochodzenia niepróchnicowego istnieją czynniki mogące utrudniać stworzenie indeksu odpowiedniego do przeprowadzenia ich oceny. Pierwszym z nich jest specyfika tych defektów, co wyraża się różnorodnością objawów klinicznych czy wzajemnym oddziaływaniem czynników etiologicznych (5).

W jednoznacznej ocenie stopnia utraty twardych tkanek zęba problemy może też stwarzać stosunek starcia zębów do wieku pacjenta. W zależności od stopnia zaawansowania ubytku, można go bowiem uznać go za patologiczny lub fizjologiczny dla danego wieku. Właściwą ocenę ułatwiłoby stworzenie standardów starcia fizjologicznego/patologicznego, przyporządkowanych do danej grupy wiekowej. Zgodnie z tym można byłoby określić zarówno stopień nasilenia defektu, jak i potrzeby lecznicze adekwatne dla danego przypadku. Bez tej standaryzacji akceptowalny poziom starcia może przy-

bierać różne wartości, w zależności od tego, czy ocenia go sam pacjent, jego lekarz, czy np. struktury opieki zdrowotnej (1, 8-10).

Kolejną dyskutowaną w piśmiennictwie kwestią jest uwzględnienie lub nieuwzględnienie kryterium stopnia odsłonięcia zębiny. Sens stosowania tego kryterium niektórzy autorzy (8, 11) podają w wątpliwość – dowodzą oni, że istnieje słaba korelacja między stopniem odsłonięcia zębiny a nasileniem starcia. Wynika to z rozkładu grubości szkliwa na powierzchni zęba. W miejscach, w których szkliwo jest najcieńsze, np. w okolicy szyjki zęba czy guzka policzkowego bliższego zębów trzonowych, nawet jego mała utrata może bowiem skutkować znacznym odsłonięciem zębiny. W badaniach rzetelności znikomy procent wyników dotyczących tego kryterium okazał się zdiagnozowany prawidłowo, niezależnie od doświadczenia naukowego badacza. Obszar odsłoniętej zębiny, w części indeksów uznawany za czynnik kluczowy, wydaje się budzić kontrowersje (8, 11). *Mulic* i wsp. za zaletę tego kryterium oceny uznają to, że funkcjonuje wiele posługujących się nim indeksów, co ułatwia porównywanie wyników. Inni autorzy stwierdzają natomiast, że np. modyfikacja *tooth wear index*, która nie rozróżnia stopnia odsłonięcia zębiny, osiąga wyższe wyniki zgodności w badaniach rzetelności (10).

W zależności od potrzeb danego badania często dokonuje się modyfikacji indeksów. Celem tych zabiegów może być dokładniejsze monitorowanie zmian wczesnych lub dostosowanie indeksu dla danej grupy wiekowej. Może to jednak skutkować trudnościami w porównywaniu wyników badań, jeśli np. uży-

skano je tylko dla dzieci i młodzieży, a brak jest badań dotyczących osób dorosłych i starszych (1, 3, 12, 13). Różnice technik wyboru próby, grup wiekowych pacjentów, a także wzajemne oddziaływanie czynników sprawczych stanowią więc czynniki komplikujące porównywanie wyników badań. Z tego względu niektóre źródła wskazują bardziej uniwersalne zastosowanie indeksów, które nie uwzględniają etiologii zmian. Przykładem może tu być najczęściej wymieniany w piśmiennictwie *tooth wear index* (TWI) wg Smitha i Knighta (3, 7, 10).

Na niekorzyść indeksów etiologicznych przemawia fakt, iż zazwyczaj uwzględniają one tylko jeden czynnik sprawczy i wymagają bardzo jasnego i precyzyjnego określenia cech morfologicznych danego defektu, a od tego, czy czynniki etiologiczne zmieniają się w czasie, może zależeć rzetelność zastosowanego instrumentu badawczego (1, 9).

Między badaczami istnieje zgodność co do podstawowych cech indeksu „idealnego”. Wymienia się tu przede wszystkim: jasność kryteriów, dopasowanie do celu badania, powtarzalność, łatwość stosowania w praktyce klinicznej. Najistotniejsza jest, jak się wydaje, przydatność indeksu do badań zarówno indywidualnych, jak i populacyjnych. Uniwersalizm instrumentu powinien wyrażać się także tym, by służył on do badań uzębienia stałego oraz mlecznego i miał zastosowanie w odniesieniu do różnych grup wiekowych pacjentów. Szczególną uwagę zwraca się na przydatność występowania indeksu w dwóch wersjach: badającej całe uzębienie oraz oceniającej jego wybrane obszary, co ma zastosowanie w badaniach przesiewowych (1-4, 6, 7).

INSTRUMENTY I STANDARDY BADAWCZE

Istotą oceny indeksu do celów badawczych i porównawczych pod kątem spełnienia wymogów jest jego standaryzacja. Rozpatrywane są tu cztery podstawowe kryteria jakościowe. Pierwszym z nich jest rzetelność (ang. reliability), która wskazuje stopień precyzji pomiaru, nie kładzie natomiast nacisku na określenie, czy dany indeks mierzy to, co powinien. Im wyższa rzetelność danego narzędzia, tym mniejszy jest błąd pomiaru (1, 3, 14). Dla celów badań medycznych istotne są głównie dwa aspekty rzetelności:

- zgodność między badaczami (ang. inter-examiner reliability) – gdy dwóch lub większa liczba badaczy przeprowadza niezależnie od siebie nawzajem badania tej samej próby,

- badanie stabilności bezwzględnej skali, inaczej stałość skali (ang. intra-examiner reliability lub tzw. test-retest) – gdy badacz sprawdza sam siebie, przeprowadzając badanie dwukrotnie (3, 14). W tym celu wybiera losowo np. 10% przypadków ze swojej próby pacjentów i po pewnym czasie ponownie przeprowadza badanie (10, 14). Niedoświadczenie badacza lub zbyt rozbudowana skala może skutkować pojawieniem się błędnych wyników (2).

Ważną kwestią jest również precyzyjne określenie warunków, w jakich mają odbywać się badania, dotyczących np.: odpowiedniego oświetlenia, pracy z ewentualnym użyciem powiększenia czy stopnia wilgotności powierzchni badanego zęba (8, 15).

Jeśli chodzi o kryterium rzetelności, to szczególną uwagę zwraca się na szkolenie badaczy, które powinno być przeprowadzane przed rozpoczęciem pracy nad indeksem. Szkolenie to może mieć charakter

dyskusji o kryteriach, w której trakcie dochodzi do wypracowania zadowalającego stopnia zgodności. W razie wątpliwości diagnostycznych, zawsze pod uwagę jest brany wynik niższy. Trening bywa przeprowadzany na podstawie badań wybranej małej próby pacjentów, a pomocne mogą się okazać fotografie wewnątrzustne czy modele diagnostyczne (2, 3, 5, 6, 8, 10, 16). Kalibracja polega na ustaleniu ilościowej relacji między wartościami uzyskanymi po zastosowaniu wybranej metody, a tzw. złotym standardem. W przypadku kalibracji badaczy, za taki wzorzec może być uznany niezależny badacz o największym doświadczeniu w badaniach klinicznych i epidemiologicznych, trenujący pozostałych. Celem kalibracji jest poznanie stopnia niedokładności pomiaru (3, 8, 10). Metodą statystyczną najczęściej stosowaną do oceny zgodności między badaczami oraz w badaniach typu „test-retest” jest współczynnik Kappa Cohena, który procentowo określa stopień zgodności uzyskanych wyników (3).

Kolejnym z podstawowych kryteriów jakościowych dla nowego narzędzia jest trafność (ang. validity). Termin ten dotyczy określenia zakresu, w jakim dany instrument mierzy to, co chce się zbadać. W prawidłowym oznaczeniu trafności podstawową rolę odgrywa pojęcie złotego standardu. W odniesieniu do złotego standardu wyznacza się tzw. trafność kryterialną. Jeśli osiąga się dodatnią korelację w porównaniu z wzorcem – oznacza to wysoką trafność kryterialną czy diagnostyczną. Ów wzorzec powinien przy tym mierzyć to zagadnienie, które jest przedmiotem zainteresowania badacza (3, 14). Jak wspomniano wcześniej, w przypadku erozji i starcia

zębów konsens co do złotego standardu nie został w pełni osiągnięty (3). Jeśli jak w powyższym przypadku, złoty standard nie istnieje, badane hipotezy należy potwierdzić doświadczalnie, za pomocą nowego, testowanego instrumentu (1).

Ze względu na częsty brak złotego standardu większego znaczenia nabiera tzw. trafność teoretyczna (ang. construct validity). Według *Brzezińskiego* jest to najważniejszy aspekt trafności, który odzwierciedla związek danego instrumentu z tzw. konstruktem teoretycznym. Aby potwierdzić trafność teoretyczną danego instrumentu, powinno się poddawać analizie obie jej części składowe, tzn. aspekt zbieżny (ang. convergent validity) oraz różnicowy (ang. discriminant validity).

Trafność zorientowana na kryterium zbieżne (ang. convergent validity) opisuje sytuację, w której obserwowane występowanie dodatniej korelacji między wynikami uzyskanymi po zastosowaniu nowego instrumentu a wynikami uznanego kryterium zewnętrznego, mierzącego obiekt podobny. *Berg-Beckhoff* i wsp. jako przykład na wykazanie trafności zbieżnej podają istnienie dodatniej korelacji między wynikami uzyskanymi za pomocą TWI a wynikami otrzymanymi podczas stosowania indeksu erozji zębów (3, 14).

Gdy między wynikami uzyskanymi po zastosowaniu nowego instrumentu oraz rezultatami użycia uznanego miernika istnieje niezbyt wysoka korelacja, a dodatkowo mierzą one nieco inny aspekt zagadnienia, mamy do czynienia z tzw. trafnością różnicową czy dyskryminacyjną (3, 14, 17). Co ciekawe, inni autorzy terminem trafności teoretycznej określają dodatnią korelację

wyników nowego instrumentu i rezultatów pomiarowych złotego standardu, nie wspominając o aspekcie trafności różnicowej (1).

Trafność treściowa, inaczej wewnętrzna (ang. content validity), jest wartością określaną subiektywnie przez grono uznanych badaczy i ma charakter opisowy. Odbywa się to przez dyskusję o tym, jakie kryteria powinny być włączone do badań i poddawane ocenie. Jeśli dany instrument charakteryzuje trafność treściowa, oznacza to, że zajmuje się on wszystkimi aspektami danego zagadnienia, bada je i czyni to we właściwy sposób, np. dobry indeks starcia nie powinien obejmować swym zasięgiem jedynie starcia powierzchni okluzyjnych, lecz także starcie innych powierzchni zębów (1, 3, 14).

Fakt, że instrument badawczy wykrywa np. starcie zębów, gdy istotnie występuje ono w jamie ustnej badanego pacjenta, dowodzi jego czułości (1, 3, 18). Kryterium to określa również, czy dany indeks jest w stanie zarejestrować zarówno zmiany uogólnione, jak i zawężone do strefy poszczególnych zębów (7).

Czułość oblicza się wg wzoru:

$$\frac{\text{Częstość odpowiedzi CP}}{\text{Częstość odpowiedzi CP} + \text{częstość odpowiedzi FN}}$$

w którym:

CP – to odpowiedź prawdziwie dodatnia (ang. correct positive), uzyskana po porównaniu z wynikami złotego standardu. Potwierdzenie

istnienia starcia, gdy rzeczywiście występuje ono w jamie ustnej.

FN – to odpowiedź fałszywie ujemna (ang. false negative), uzyskana po porównaniu z wynikami złotego standardu. Wykluczenie istnienia starcia, podczas gdy jest ono obiektywnie obecne w jamie ustnej (1, 3, 18).

Swoistość (ang. specificity) – określa stopień niewykrycia defektu przez indeks, jeśli defektu tego istotnie się nie stwierdza (1, 3, 18). Mierzy się ją za pomocą zależności:

$$\frac{\text{Częstość odpowiedzi CN}}{\text{Częstość odpowiedzi CN} + \text{częstość odpowiedzi FP}}$$

w którym:

CN – odpowiedź prawdziwie ujemna (ang. correct negative), uzyskana po porównaniu z wynikami wzorca. Niewykrycie defektu, gdy istotnie go nie ma.

FP – odpowiedź fałszywie dodatnia (ang. false positive), uzyskana po porównaniu z wynikami wzorca. Stwierdzenie istnienia defektu, podczas gdy obiektywnie nie występuje on w jamie ustnej.

W celu pełnej oceny instrumentu badawczego powinno się dokonać sprawdzenia wszystkich wymienionych kryteriów jakościowych (1, 3).

Podczas dyskusji na temat wyboru kryteriów dla indeksów rejestrujących erozję czy starcie zębów zachodzi także potrzeba ustalenia, które z obecnych w jamie ustnej zębów mają być wykluczone z badania. Dyskwalifikowane mogą być

np. zęby z zaawansowanymi zmianami próchnicowymi, rozległymi wypełnieniami, z obecnością zamków aparatów ortodontycznych lub złamań na powierzchni zęba (10). Niektóre z indeksów erozji pomijają także zęby z widocznymi ubytkami klinowymi i oznakami atrycji (2). Omówienia może również wymagać kwestia ewentualnego wpływu na wyniki badań czynników socjoekonomicznych, kulturowych, geograficznych, zdrowia ogólnego (8, 10). Oprócz badania klinicznego, dla jakości oceny opisywanych defektów może się okazać właściwe zastosowanie technik pomocniczych, takich jak np. fotografie, kwestionariusze czy modele diagnostyczne przydatne w analizie retrospektywnej starcia (2, 3, 5, 6, 8, 10, 16). Przy próbie zachowania wszystkich kryteriów jakościowych indeksów dąży się do upraszczania ich wersji, co pozwala na bardziej ekonomiczne i mniej czasochłonne ich stosowanie, a także ułatwia trening i kalibrację badaczy (7).

Stworzenie indeksu w pełni odzwierciedlającego naturę i rozległość defektów, który byłby łatwy do zastosowania w codziennej praktyce klinicznej, dokładnie wystandaryzowany i uniwersalny pozostaje nadal w sferze badań. Wyzwanie to jednak jest przez kompetentne gremia podejmowane, łącznie z propozycjami indeksów określających wskazania terapeutyczne dla opisywanych zmian (1, 2). ■

PIŚMIENNICTWO

1. Young A. i wsp.: Current erosion indices – flawed or valid? Summary. Clin. Oral Invest., 2008, 12, Suppl. 1, 59-63.

2. Mulic A. i wsp.: Reliability of two clinical scoring systems for dental erosive wear. Caries Res., 2010, 44, 3, 294-299.

INSTRUMENTY I STANDARDY BADAWCZE

3. *Berg-Beckhoff G., Kutchmann M., Bardehle D.*: Methodological considerations concerning the development of oral dental erosion indexes: literature survey, validity and reliability. *Clin. Oral Invest.*, 2008, 12, Suppl. 1, 51-58.
4. *Bardsley P.F.*: The evolution of tooth wear indices. *Clin. Oral Invest.*, 2008, 12, Suppl. 1, 15-19.
5. *Hooper S.M., Meredith N., Jagger D.C.*: The development of a new index for measurement of incisal/occlusal tooth wear. *J. Oral Rehabil.*, 2004, 31, 3, 206-212.
6. *Bartlett D.W.*: Retrospective long-term monitoring of tooth wear using study models. *Br. Dent. J.*, 2003, 194, 4, 211-213.
7. *Steele J.G., Walls A.W.*: Using partial recording to assess tooth wear in older adults. *Community Dent. Oral Epidemiol.*, 2000, 28, 1, 18-25.
8. *Ganss C., Klimek J., Giese K.*: Dental erosion in children and adolescents – a cross-sectional and longitudinal investigation using study models. *Community Dent. Oral Epidemiol.*, 2001, 29, 4, 264-271.
9. *Richards L.C., Kaidonis J.A., Townsend G.C.*: A model for the prediction of tooth wear in individuals. *Aust. Dent. J.*, 2003, 48, 4, 259-262.
10. *de Carvalho Sales-Peres S.H.* i wsp.: Prevalence of dental wear among 12-year-old Brazilian adolescents using a modification of tooth wear index. *Public Health*, 2008, 122, 9, 942-948.
11. *Holbrook W.P., Ganss C.*: Is diagnosing exposed dentine a suitable tool for grading erosive loss? *Clin. Oral Invest.*, 2008, 12, Suppl. 1, 33-39.
12. *Ganss C.*: How valid are current diagnostic criteria for dental erosion? *Clin. Oral Invest.*, 2008, 12, Suppl. 1, 41-49.
13. *Fares J.*: A new Index of Tooth Wear. *Caries Res.*, 2009, 43, 2, 119-125.
14. *Brzeziński J.*: Metodologia badań psychologicznych. Wyd. Nauk. PWN, Warszawa 1999.
15. *Woźniak J.* i wsp.: Powtarzalność pomiarów fluorescencji za pomocą aparatu Diagnodent na powierzchniach zużywających zębów trzonowych w warunkach *in vitro*. *Dent. Med. Probl.*, 2008, 45, 3, 255-259.
16. *Larsen I.B.* i wsp.: A clinical index for evaluating and monitoring dental erosion. *Community Dent. Oral Epidemiol.*, 2000, 28, 3, 211-217.
17. *Słupczyńska-Kossobudzka E., Wójtowicz S., Boguszewska L.*: Własności psychometryczne kwestionariusza zachowań uciążliwych. *Post. Psychiatr. Neurol.*, 2003, 12, 3, 277-286.
18. *Jędrzychowski W.*: Zasady planowania i prowadzenia badań naukowych w medycynie. Wyd. I. Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2004.