

Development of innovative training solutions in the field of functional evaluation aimed at updating of the curricula of health sciences schools



Moduł Biomechanika chodu

Jednostka dydaktyczna C: Jak oceniać chód?

C.3: Jakie są zalety stosowania technik instrumentalnych w porównaniu ze skalą i badaniem fizycznym do oceny chodu?



Spis treści:

1. WPROWADZENIE I CELE	2
1. CECHY I WŁAŚCIWOŚCI NARZĘDZI DO OCENY CHODU: PORÓWNANIE DOSTĘPNYCH TECHNIK	3
2.2 Wymagania dotyczące sprzętu	5
2.3 Obiektywność wyników i analizy statystycznej	8
2.4 Ważność.....	11
2.5 Niezawodność	14
2.6 Wrażliwość na zmiany i reaktywność.....	20
2.7 Efekt podłogi i sufitu	24
3.KLUCZOWE IDEE	26
4.BIBLIOGRAFIA	27

1. Wprowadzenie i cele

Zarówno ocena chodu poprzez 1) obserwację, 2) skale kliniczne, testy i kwestionariusze, jak i 3) ocena przy użyciu technik instrumentalnych są przydatne do poznania wzorca chodu danej osoby, ale istnieją zasadnicze różnice, które należy wziąć pod uwagę przy wyborze jednej lub drugiej techniki oceny lub kombinacji obu.

W tej jednostce dokonamy przeglądu cech metodologicznych i właściwości statystycznych technik dostępnych do oceny chodu człowieka, omówionych w Jednostce C.1 modułu Biomechanika Chodu.

W oparciu o to, proponujemy następujące cele:

1. Dokonanie przeglądu zalet i wad metodologii oceny chodu człowieka.
2. Poznanie właściwości statystycznych dostępnych metodologii oceny chodu.
3. Stworzenie wiedzy technicznej, która pozwoli pracownikom służby zdrowia na wybór najbardziej odpowiedniej techniki oceny chodu dla ich kontekstu klinicznego lub badawczego.

1. Cechy i właściwości narzędzi do oceny chodu: porównanie dostępnych technik

2.1 Użyteczność

Użyteczność odnosi się do łatwości, z jaką ludzie mogą używać danego narzędzia, by osiągnąć określony cel. Mając do czynienia z praktyką zawodową o dowolnym profilu zdrowotnym i w dowolnym obszarze, przed wyborem techniki oceny (lub leczenia), specjaliści będą zadawać pytania takie jak: czy jest łatwa w użyciu? Czy zajmuje to dużo czasu? Czy jest to wykonalne do zastosowania w moim miejscu pracy? W tym sensie, dostępne techniki oceny chodu człowieka (omówione w poprzednich rozdziałach) mają różne odpowiedzi na pytania postawione powyżej.

Jeśli chodzi o łatwość użycia, techniki instrumentalne wymagają kilku kroków ujętych w ścisły protokół oceny, który musi być przestrzegany, aby zapewnić, że porównanie pomiarów z danymi własnymi pacjenta (na przykład, gdy kilka ocen jest przeprowadzanych w okresie czasu) lub z danymi innych uczestników jest ważne. To sprawia, że każda technika instrumentalna jest bardziej złożona w użyciu niż skala, kwestionariusz lub test kliniczny (Tabela 1). W protokole pomiarowym z zastosowaniem techniki instrumentalnej prawdopodobnie najbardziej złożonym etapem jest instrumentacja osoby badanej, jeśli jest konieczna, oraz obróbka danych po wykonaniu pomiaru. Dlatego też, biorąc pod uwagę powyższe, instrumentalne techniki oceny chodu wymagają uwzględnienia bardzo istotnego czynnika: czasu (Tabela 1). Jednakże, wraz z postępem dostępnych technologii, w obrębie grupy technik instrumentalnych do pomiaru tych samych parametrów, możemy znaleźć różne scenariusze.

Przykład

Biomechaniczna instrumentalna ocena parametrów przestrzenno-czasowych chodu

Wykorzystując system fotogrametrii i oprzyrządowany chodnik, możliwe jest dokonanie pomiaru parametrów przestrzenno-czasowych charakteryzujących chód człowieka. Parametry te to przede wszystkim prędkość chodu, długość kroku i rozkroku, kadencja, czas podwójnego podparcia oraz czas fazy podporu i wymachu w cyklu chodu. Mimo, że za pomocą obu technik możemy osiągnąć ten sam wynik, to czas potrzebny na wykonanie każdej z nich różni się znacząco.

W przypadku oprzyrządowanego chodnika (z czujnikami nacisku) nie jest konieczne oprzyrządowanie osoby badanej, ponieważ osoba oceniana powinna jedynie przejść przez korytarz w obuwiu lub bez. Jeśli oprogramowanie zarządzające sprzętem oblicza parametry na podstawie zarejestrowanych ciśnień na stopę, nie będzie konieczne wykonywanie dalszych zabiegów w celu zmierzenia danych. W związku z tym, osoba oceniająca nie potrzebuje nadmiernego szkolenia lub wiedzy, aby korzystać z tego narzędzia.

Z drugiej strony, w przypadku systemu fotogrametrycznego, oprzyrządowanie badanego jest zawsze konieczne. To oprzyrządowanie składa się z biomechanicznego modelu punktów

orientacyjnych, które są umieszczone w ciele, aby reprezentować segmenty zaangażowane w ruch, który ma być oceniony. W zależności od złożoności modelu, czas oprzyrządowania będzie dłuższy lub krótszy. W każdym przypadku, skóra w miejscu mocowania markera musi być zawsze oczyszczona i ogolona. Z drugiej strony, oprogramowanie zarządzające systemami fotogrametrycznymi zwykle podaje jako wyniki współrzędne każdego markera w każdej osi ruchu i w każdym czasie rejestracji obrazu (na przykład 50 pozycji na sekundę, jeśli częstotliwość rejestracji wynosi 50 Hz). Jeśli oprogramowanie nie oferuje protokołu z zamkniętym i stałym modelem i wcześniej nie oblicza różnych parametrów, które można uzyskać z pomiaru, obliczenie wyników zawsze będzie musiało być wykonane po wyodrębnieniu współrzędnych każdego punktu orientacyjnego z modelu biomechanicznego. Wszystko to wymaga, aby oceniający posiadał głębszą wiedzę i ćwiczył prawidłowe umiejscowienie markerów.

Instrumented walkway

Photogrammetry system



Subject instrumentation:	NO	YES
Data processing after measurement:	NO	YES
Training evaluator :	NO	YES
Approximate assessment time:	5 minutes	1 hour

Z drugiej strony, obserwacyjna analiza chodu lub ocena za pomocą skal klinicznych wymaga minimalnej ilości czasu, a niezbędne szkolenie do oceny chodu jest związane z praktyką instrukcji samej skali lub testu. To właśnie z powyższych względów zastosowanie każdej z technik pomiarowych ma większą przydatność w różnych dziedzinach. Wykorzystanie oceny chodu za pomocą skal lub testów klinicznych czy analizy chodu metodą eye-naked jest łatwiejsze do zastosowania w kontekście codziennych konsultacji klinicznych, gdzie pacjenci występują z różnymi patologiami prowadzącymi do zaburzeń chodu. Z drugiej strony, techniki instrumentalne są zazwyczaj wykorzystywane w projektach badawczych. Mimo to, w kontekście badań naukowych, skale i zwalidowane testy kliniczne w różnych populacjach i językach są szeroko stosowane.

Tabela 1 - Użyteczność narzędzi do oceny chodu

Charakterystyka	Analiza obserwacyjna chodu	westionariusz, skale i testy kliniczne	Techniki instrumentalne
Koszt czasu	+	+	+ / ++/ +++ (w zależności od zastosowanego systemu)
Szkolenie oceniającego	+	+	++ / +++ (w zależności od zastosowanego systemu)
Kontekst użycia	Kliniczna	Kliniczne i naukowe badania	Badania
Użyteczność	+	++	+++

2.2 Wymagania dotyczące sprzętu

Podczas gdy ocena chodu za pomocą skali lub kwestionariusza nie wymaga niczego więcej niż odpowiedniego formularza i długopisu, testy kliniczne mogą wymagać pewnej ilości materiałów, aby zmontować odpowiedni zestaw. Niektóre testy mogą wymagać krzesła, taśmy do oznaczania podłoża i stopera, podczas gdy inne mogą wymagać większej ilości materiałów (patrz przykład poniżej). W każdym przypadku materiały te są zazwyczaj niedrogie (Tabela 2) lub dostępne dla każdego pracownika służby zdrowia.

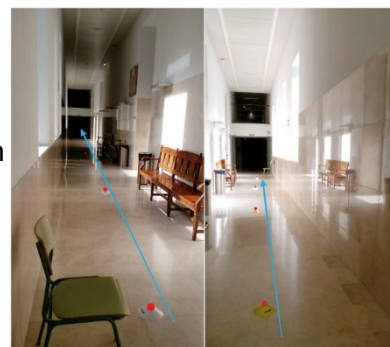
Przykład

Ile materiałów potrzebuję do wykonania 6-minutowego testu chodu?

Test 6-minutowego chodzenia (6MWT) jest również stosowany jako jednorazowy miernik stanu funkcjonalnego pacjentów, a także jako predyktor zachorowalności i śmiertelności. Test 6MWT jest praktycznym, prostym testem, który wymaga 100-metrowego korytarza, ale nie wymaga sprzętu do ćwiczeń ani zaawansowanego szkolenia dla techników. Test ten mierzy odległość, jaką pacjent może szybko pokonać po płaskiej, twardej powierzchni w ciągu 6 minut (6MWD). Ocenia on globalne i zintegrowane reakcje wszystkich systemów zaangażowanych podczas wysiłku, w tym układu płucnego i sercowo-naczyniowego, krążenia systemowego, krążenia obwodowego, krwi, jednostek nerwowo-mięśniowych i metabolizmu mięśni. Samoczynna próba wysiłkowa 6MWT ocenia submaksymalny poziom wydolności funkcjonalnej.

Wymagane wyposażenie to:

1. Zegar odliczający czas (lub stoper)
2. Mechaniczny licznik okrążeń
3. Dwa małe pachołki do oznaczenia punktów zwrotnych
4. Krzesło, które można łatwo przemieszczać wzdłuż trasy marszu
5. Arkusze na tablicy ogłoszeń (formularz rejestracyjny)
6. Taśma klejąca lub kolorowe naklejki do oznaczenia miejsca zatrzymania pacjenta w czasie 6 minut
7. Skala Borga
8. Pulsoksymetr
9. Sfigmomanometr i stetoskop
10. Telefon
11. Źródło tlenu
12. Automatyczny defibrylator elektroniczny



Ustawienie testu 6-minutowego chodzenia

Jednakże, jeśli chodzi o ocenę za pomocą technik instrumentalnych, wymagany materiał jest nie tylko bardziej ilościowy, ale również wysoce specjalistyczny. Ogólnie rzecz biorąc, sprzęt pomiarowy do rejestracji wzorca chodu lub określonej cechy chodu składa się z następujących części:

- **Czujniki lub/i sprzęt pomiarowy** odnosi się do urządzeń, które składają się na sam sprzęt pomiarowy. Na przykład:

o Platforma dynamometryczna składa się ze sztywnej, płaskiej i statycznej powierzchni zainstalowanej na podłożu oraz czujników lub przetworników, które są umieszczone w dolnej części platformy, rozmieszczonych w taki sposób, aby były w stanie rejestrować siły wywierane na górną sztywną powierzchnię.

o Sprzęt do elektromiografii składa się z elektrod (czujników), które zbierają aktywność elektryczną w mięśniu, poprzez włożenie ich lub przez skórę, która pokrywa mięsień. System obejmuje również wzmacniacz, który umożliwia przesyłanie sygnałów elektrycznych z mięśnia do monitora, zazwyczaj bezprzewodowo, oraz okablowanie pomiędzy czujnikami a wzmacniaczem.

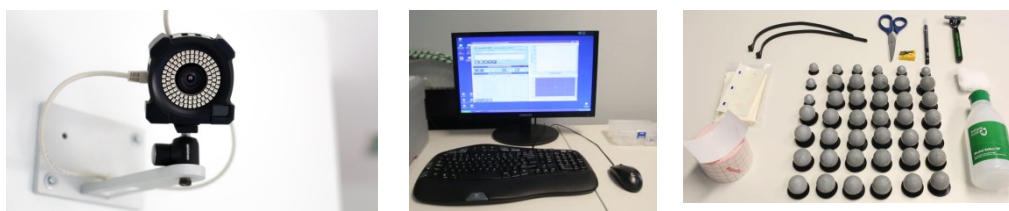
- **Oprogramowanie i komputer:** odnosi się do oprogramowania, które umożliwia zarządzanie urządzeniami rejestrującymi i czujnikami, przechowywanie zarejestrowanych danych oraz komputer, na którym oprogramowanie jest używane.
- **Materiały eksploatacyjne:** odnosi się do wszelkiego rodzaju akcesoriów wymaganych do przeprowadzenia pomiaru chodu. Na przykład:

o W systemie elektromiografii niezbędne akcesoria to te, które posłużą do przygotowania skóry, na której zostanie umieszczona elektroda, a tym samym do zmniejszenia impedancji pomiędzy elektrodą a skórą: golarka, alkohol, bawełna, drobny papier ścierny do usunięcia martwego naskórka. Dodatkowo potrzebny

będzie żel przewodzący, który również pomoże zmniejszyć impedancję skóra-elektroda.

Przykład

Ile materiałów jest potrzebnych do przeprowadzenia oceny chodu za pomocą systemu fotogrametrycznego?



Rysunek 1 - System fotogrametryczny i jego komponenty

Poza kamerami, oprogramowaniem i komputerem, w systemie fotogrametrycznym niezbędne będzie posiadanie zestawu punktów orientacyjnych, które będą reprezentować model biomechaniczny. Dodatkowo, aby były one dobrze przylegające do skóry, potrzebny będzie system klejów, elementy do oczyszczania skóry i zapobiegania ich odpadaniu w trakcie badania (golarka, alkohol, bawełna) oraz odzież niezbędna do wykonania badania.

Jak widać, materiały i akcesoria technik instrumentalnych do oceny chodu nie są zwykle dostępne w placówkach medycznych. Koszt sprzętu i oprogramowania może być bardzo zróżnicowany w zależności od instrumentu oceny i możliwości oprogramowania (tab. 2). Na przykład niektóre oprogramowanie może oferować dane normatywne, z którymi można porównać ocenę dokonaną na pacjencie o określonej płci i wieku, co zwiększa jego cenę. Sprzęt bez żadnego oprzyrządowania, taki jak chodnik z czujnikami ciśnienia i odpowiednim oprogramowaniem, może kosztować około 40 tys. euro w zależności od długości korytarza, podczas gdy sprzęt fotogrametryczny z kamerami i oprogramowaniem może z łatwością przekroczyć 100 tys. euro.

Tabela 2 - Wymagania dotyczące narzędzi do oceny chodu

Charakterystyka	Obserwacyjna analiza chodu	Kwestionariusz, skale i testy kliniczne	Techniki instrumentalne
Sprzęt	+	+	+++
Materiały eksploatacyjne	-	+	++
Koszt ekonomiczny	+	+	+++

2.3 Obiektywność wyników i analizy statystycznej

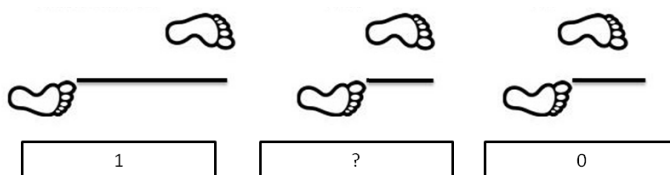
Obiektywne dane oceny chodu odnoszą się do danych uzyskanych bez interpretacji osoby oceniającej (tj. bezpośrednia ocena jednego lub więcej wymiarów wzorca chodu), podczas gdy w danych subiektywnych wynik podlega interpretacji, percepcji lub opinii osoby oceniającej. W dostępnych narzędziach do oceny chodu znajdujemy oba rodzaje wyników. Podczas gdy za pomocą instrumentalnych narzędzi oceny możemy uzyskać obiektywne dane pomiarowe, analiza chodu poprzez obserwację, kwestionariusze lub skale kliniczne dostarczy informacji subiektywnych. Z tego powodu dobrze jest, aby pracownik służby zdrowia wiedział przed jej zastosowaniem, że skale oceny są bardzo przydatnymi narzędziami do jakościowego określenia ogólnego stanu zdrowia, jednak podlegają one w pewnym stopniu subiektywizmowi zarówno pacjenta (ocena na podstawie deklaracji własnych), jak i osoby oceniającej (ocena na podstawie obserwacji). Ponadto, czasami są one opracowywane w bardzo specyficznych kontekstach (populacje, regiony, kraje, itp.), z bardzo ważnymi demograficznymi i kulturowymi uprzedzeniami, które utrudniają lub uniemożliwiają ekstrapolację na inne społeczności. Z drugiej strony, główną zaletą ocen opartych na raportach własnych jest to, że dają one obraz tego, co dana osoba naprawdę przeżywa i postrzega. Ponadto, metoda ta jest tania, szybka i wykonalna.

Przykład

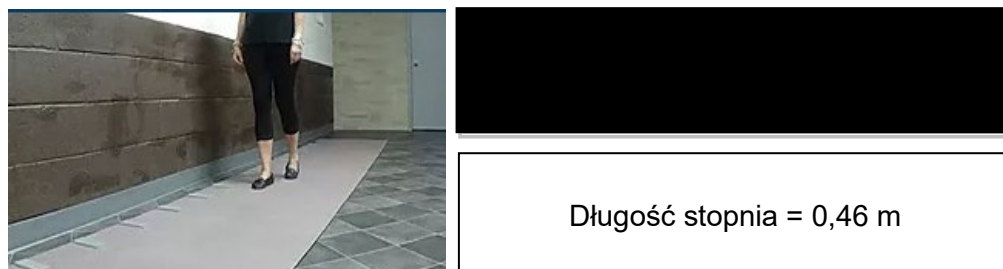
Subiektywna i obiektywna miara długości kroku

Wiele skal oceny chodu posiada pewne pozycje związane z długością kroku. Jedną z nich jest Tinetti Mobility Test (TMT), w której w części dotyczącej chodu pyta się, czy pacjent jest w stanie przekładać jedną stopę nad drugą podczas stawiania kroku. Kiedy mierzymy długość kroku u osoby bez patologii, odpowiedź może być wyraźnie widoczna i obserwator niewątpliwie zwróci uwagę, że prawa stopa przewyższa lewą (lub odwrotnie). Jeśli jednak ocenie poddawany jest pacjent z chorobą Parkinsona, u którego jedną z cech choroby jest chodzenie małymi krokami, może się zdarzyć, że odpowiedź na pozycję TMT nie będzie tak jednoznaczna i dwóch badaczy zaobserwuje różne rzeczy. Z drugiej strony, urządzenie pomiarowe, takie jak chodnik z oprzyrządowaniem, jest w stanie zmierzyć długość kroku poprzez wykrywanie nacisku na podłoże, gdy pacjent stawia stopę na podłożu. W przypadku tego typu przyrządu, na odległość uzyskaną z długości kroku nie ma wpływu obserwacja oceniającego.

Step length and height	
The left foot does not pass right stance foot	0
The left foot passes right stance foot with step	1



Rysunek 2 - Długość i wysokość kroku z Testu Mobilności Tinetti, sekcja Chód. Pokazane są trzy przykłady długości kroku, w których oceniający musi zaznaczyć 1 lub 0 zgodnie z kryteriami ustalonymi przez skalę. W przykładzie środka może pojawić się wątpliwość, jaką numerację przypisać.



Rysunek 3 - Ocena długości kroku za pomocą oprzyrządowanego przejścia (GAITrite)

Główną zaletą stosowania obiektywnych wyników mierzonych technikami instrumentalnymi jest to, że pozwalają one, między innymi, na porównywanie danych z innymi danymi pochodzącymi od tego samego pacjenta lub z wynikami uzyskanymi od innych osób. Jednakże, badacz lub profesjonalny oceniający nie powinien zapominać, że porównanie obiektywnych danych pomiędzy osobami powinno być normalizowane przez najbardziej wpływową cechę na te dane. Na przykład, na długość kroku wpływa również wzrost osoby badanej, a zatem długość kończyn dolnych, dlatego obiektywne dane dotyczące długości kroku powinny być podzielone przez wzrost lub długość nóg. Podobnie będzie w przypadku takich zmiennych jak siły reakcji i waga badanych.

Mimo że skale lub kwestionariusze oraz narzędzia instrumentalne nie dostarczają tego samego rodzaju informacji (subiektywne versus obiektywne), subiektywne pomiary mogą być wysoce skorelowane z obiektywnymi. Jest to kolejna wartość dodana do skal oceny stosowanych w warunkach klinicznych. Skala lub kwestionariusz, który jest wysoce skorelowany z wynikami oceny za pomocą techniki instrumentalnej, będzie ważnym narzędziem do pomiaru populacji w dużych próbach.

Przykład



Journal List > Age Ageing > PMC4476851

OXFORD JOURNALS Age and Ageing

Age Ageing. 2015 Jul; 44(4): 691–694. PMID: PMC4476851
Published online 2015 May 27. doi: [10.1093/ageing/afv062](https://doi.org/10.1093/ageing/afv062) PMID: [26018999](https://pubmed.ncbi.nlm.nih.gov/26018999/)

A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study

Paul Innerd,¹ Michael Catt,¹ Joanna Collerton,² Karen Davies,² Michael Trenell,³ Thomas B. L. Kirkwood,⁴ and Carol Jagger⁴

W badaniach epidemiologicznych aktywność fizyczna jest powszechnie oceniana za pomocą kwestionariuszy aktywności fizycznej ze względu na ich praktyczność i niski koszt. Ocena dużych prób uczestników za pomocą techniki instrumentalnej byłaby niezwykle kosztowna i skomplikowana. Dlatego też istnieją badania, które mają na celu określenie korelacji kwestionariuszy własnej aktywności fizycznej z obiektywnymi miarami, takimi jak aktywność fizyczna rejestrowana za pomocą akcelerometrów stosowanych w małej próbie populacji docelowej przez jeden tydzień. Celem tego typu badań jest więc ocena ważności testu subiektywnego jako narzędzia do oceny aktywności fizycznej dla dużych prób uczestników.

Czy dane subiektywne mogą być analizowane statystycznie? Odpowiedź brzmi: tak, zarówno subiektywne informacje uzyskane za pomocą skal lub kwestionariuszy, jak i obiektywne informacje uzyskane za pomocą techniki instrumentalnej, mogą być analizowane statystycznie. Jest oczywiste, że np. prędkość chodu jest zmienną ilościową, tzn. taką, która posiada wielkości liczbowe jako modalności, na których możemy wykonywać operacje arytmetyczne (tab. 3).

Tabela 3 - Obiektywność wyników i analizy statystycznej

Charakterystyka	Obserwacyjna analiza chodu	Kwestionariusz, skale i testy kliniczne	Techniki instrumentalne
Obiektywność	Informacje subiektywne	Informacje subiektywne	Informacje obiektywne
Analiza statystyczna	Nie	Tak	Tak
Rodzaj zmiennej w analizie statystycznej	-	Jakościowe, półilościowe i ilościowe (w zależności od odpowiedzi do analizy)	Ilościowe

Z drugiej strony, na podstawie kwestionariusza lub skali, możemy analizować jego wynik końcowy. Jeśli ten wynik końcowy jest liczbą, to będzie on analizowany statystycznie jako zmienna ilościowa, bo taką wprowadzimy do programu statystycznego (liczba), ale nadal jest to informacja subiektywna. Jeśli natomiast analizie poddamy każde pytanie ze skali lub kwestionariusza, możemy uzyskać różne zmienne statystyczne (tabela 3), w zależności od rodzaju odpowiedzi, jakie uda nam się znaleźć:

1. Dynamiczna Skala Chodu Parkinsona zawiera pytania, w których oceniający musi przypisać ocenę od 0 do 5 w odniesieniu do sprawności chodu pacjenta. Na przykład, w pozycji 1 odnoszącej się do tego, jak pacjent idzie siedem metrów do przodu, oceniający może ocenić tę cechę od normalnej (0 punktów) do "niezdolny do zainicjowania kroku do przodu lub nieudany" (5

punktów). Mimo, że jest to nadal kategorię modalność w pytaniu, te możliwe odpowiedzi mają swoją kolejność: jeden jest gorszy niż zero, dwa są gorsze niż jeden, trzy są gorsze niż dwa, i tak dalej. Jest to zmienna półilościowa.

Dynamic Parkinson Gait Scale (DYPAGS)	
	Score
1. Walking 7 m forwards	
Normal	0
Subtle start hesitation (<1 s) or slow gait or increased double-support time	1
Start hesitation >1 s or destination hesitation or impaired feet clearance	2
Block or accelerated short steps	3
Unable to perform the entire distance or near fall	4
Unable to initiate a step forward or fall	5

2. Test Mobilności Tinetti zawiera pytania, w których obserwator musi zweryfikować cechę chodu u pacjenta i przyznać 0 punktów, jeśli to, co obserwuje jest negatywne lub zmieniona odpowiedź lub 1 punkt, jeśli cecha chodu oceniana jest przedstawiona jako pozytywna lub normalna. Innymi słowy, tego typu pytania lub skale mają charakter odpowiedzi "tak/nie". Przy analizie statystycznej, uważalibyśmy to za jakościową, kategorię modalność z dwoma modalnościami: 0 = cecha chodu zmienionego i 1 = cecha chodu prawidłowego.

Walking time	Heels apart	= 0
	Heels almost touching while walking	= 1

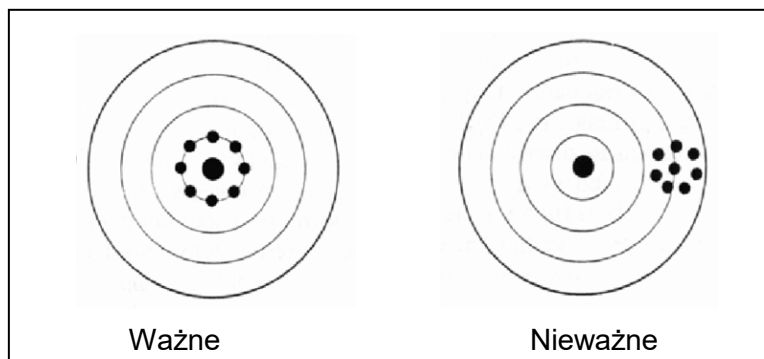
Wreszcie, gdy obserwator przeprowadza analizę chodu pacjenta gołym okiem i uzyskuje zestaw pomysłów lub cech charakterystycznych dotyczących wzorca chodu, który obserwuje, informacja ta jest subiektywna, ale ponieważ nie jest standaryzowana, nie może być analizowana przy użyciu konwencjonalnych metod statystycznych (Tabela 3).

2.4 Ważność

Instrumenty, zanim zostaną uznane za odpowiednie, muszą oferować dokładne, ważne i możliwe do zinterpretowania dane do oceny stanu zdrowia populacji. Co więcej, środki te powinny dostarczać naukowo solidnych wyników. Skuteczność wyników tych działań wynika z rzetelności i ważności instrumentów.

Ważność odnosi się do faktu, że narzędzie mierzy dokładnie to, co proponuje mierzyć lub, innymi słowy, ważność odnosi się do dokładności pomiaru (Ryc. 4). Ważność nie jest cechą instrumentu i musi być określona w odniesieniu do konkretnej sprawy, gdy odnosi się do zdefiniowanej populacji. Właściwości pomiaru - ważność i rzetelność - nie są całkowicie niezależne. Narzędzie, które nie jest rzetelne, nie może być ważne; natomiast narzędzie

rzetelne może być niekiedy nieważne. Tak więc wysoka rzetelność nie gwarantuje ważności narzędzia..



Rysunek 4 - W celu po lewej stronie, strzały trafiały w miejsce, w które miały trafić i były spójne, dokładnie w środek celu. W celu po prawej stronie, strzały były niezawodne, trafiając w ten sam punkt; jednak żaden nie trafił w środek celu, nie będąc uznany za ważny (Obraz z De Souza A. et al. 2017)

Wiarygodność narzędzi oceny wymaga kilku źródeł dowodów, aby zbudować argument, że narzędzie mierzy to, co powinno mierzyć. Określenie ważności może być postrzegane jako konstruowanie argumentu opartego na dowodach dotyczących tego, jak dobrze narzędzie mierzy to, co powinno robić. Dowody mogą być zebrane w celu wsparcia lub braku wsparcia dla konkretnego zastosowania narzędzia oceny. Dlatego też w bibliografii można znaleźć kilka badań, których celem jest zbadanie ważności narzędzia do pomiaru chodu w określonej populacji docelowej (Gail M. Sullivan 2011).

Jaka jest procedura pomiaru ważności narzędzia?

Technika lub narzędzie, którego ważność ma być badana, musi być porównana ze "złotym standardem" narzędzia, którego ważność została już szeroko przetestowana do pomiaru określonego wyniku (ważność równoległa). W tego typu badaniach, celem jest odpowiedź na pytanie: czy narzędzie A mierzy tak samo dokładnie jak narzędzie B chód człowieka? Zwykle analizuje się to za pomocą współczynnika korelacji Pearsona lub Spearmana (r). Uważa się, że jest ona doskonała, gdy współczynnik korelacji pomiędzy narzędziem pomiarowym a wzorcem odniesienia wynosi powyżej 0,6, adekwatna pomiędzy 0,59 a 0,31, a słaba poniżej 0,3.

Przykład 1: ważność skali

Diane M. Wrisley i Neeraj A. Kumar (2010) określili ważność funkcjonalnej oceny chodu (Functional Gait Assessment - FGA) do klasyfikacji ryzyka upadku i przewidywania niewyjaśnionych upadków u starszych osób mieszkających w społeczności. W tym celu trzydziestu pięciu starszych dorosłych w wieku od 60 do 90 lat wypełniło Skalę Zachowania Równowagi (ABC), Skalę Równowagi Berga (BBS), Dynamiczny Wskaźnik Chodu (DGI), Test Wstawania i Chodzenia w Czasie (TUG) oraz narzędzie poddane walidacji (FGA). Współczynniki korelacji Spearmana zostały użyte do określenia ważności współbieżnej pomiędzy ABC, BBS, TUG, DGI i FGA. FGA korelowało z ABC ($r = .053$, $P < .001$), BBS ($r =$

.84, $P < .001$) i TUG ($r = .84$, $P < .001$). Autorzy stwierdzili, że FGA z wynikiem odcięcia 22/30 jest skuteczny w klasyfikacji ryzyka upadku u starszych osób i przewidywaniu niewyjaśnionych upadków u starszych osób mieszkających w społeczności.

Przykład 2: ważność trójwymiarowej analizy chodu na bieżni

Pinto R. i wsp. (2020) określili ważność trójwymiarowej (3D) biomechaniki chodu uzyskanej z systemów opartych na bieżni. W tym celu badacze zbadali równoczesną ważność poprzez oszacowanie powiązań pomiędzy pomiarami opartymi na bieżni i naziemnymi (złoty standard) u uczestników z chorobą zwyrodnieniową stawu kolanowego. Prędkość chodu na bieżni została dopasowana do samodzielnie wybranej prędkości na powierzchni ziemi. Ustawienie markerów, kąt kolana i obliczenia momentu były spójne dla obu systemów. Zmienne obliczone na podstawie kątów i momentów kolan podczas chodu oceniano przy użyciu korelacji Pearsona (r) oraz innych parametrów statystycznych. Wyniki wykazały, że korelacje Pearsona pomiędzy bieżnią i systemem naziemnym wahały się od 0,56 do 0,97. Pomimo wysokiej korelacji, istniały istotne różnice w momentach, co podkreśla, że nie mogą być one stosowane zamiennie. Sugeruje to, że kąty i momenty kolana w płaszczyźnie czołowej i strzałkowej u pacjentów z chorobą zwyrodnieniową stawu kolanowego oceniane za pomocą systemu opartego na bieżni ruchomej są prawidłowe.

Teraz, kiedy już zdefiniowaliśmy ważność i jak ona działa, warto zadać sobie pytanie: jakie rodzaje narzędzi mają największą ważność do pomiaru chodu lub konkretnej cechy chodu? Skale / testy kliniczne czy techniki instrumentalne? Odpowiedź jest jasna, instrumentalne techniki pomiarowe są bardziej wiarygodne do pomiaru chodu niż skale czy testy kliniczne, ponieważ są to bardziej precyzyjne narzędzia do pomiaru określonej zmiennej chodu. Warto wspomnieć, że nawet wśród technik instrumentalnych do analizy chodu, są takie, które są bardziej precyzyjne od innych. Te różnice w precyzji są związane z błędem pomiarowym sprzętu, implikowanym przez samą metodologię. Rysunek 5 przedstawia porównanie dostępnych technik instrumentalnych do pomiaru parametrów przestrzenno-czasowych chodu, takich jak prędkość, długość kroku, kadencja, czas trwania cyklu chodu i fazy chodu.

Devices	Precision	Cost
Chronometer	+	+
Pedometer	+	+
GPS	++	++
Radar speed	+++	++++
Cross line detector	+++	++
Inertial measurement unit	++	+++
Footswitch	+++	++
Instrumented walkway	+++	++++
Optoelectronic cameras	++++	+++++

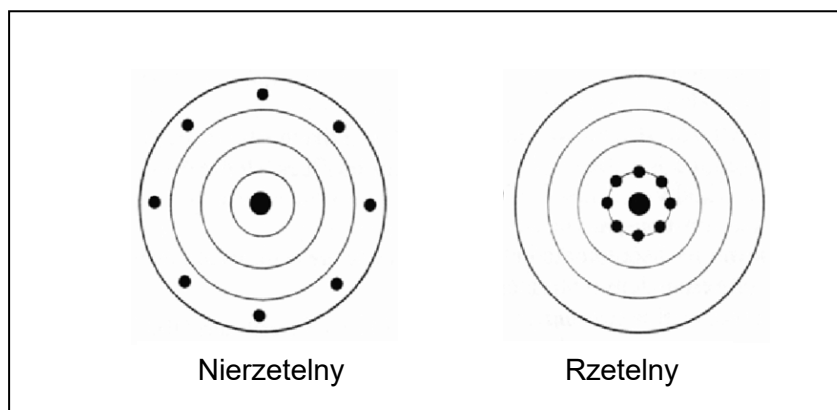
Ryc. 5 - Porównanie powszechnie stosowanych technologii do pomiaru parametrów spatio-temporalnych chodu (Moissenet F. i Armand S. 2016). Dla każdej techniki instrumentalnej podano stopień precyzji i koszt techniki.

2.5 Niezawodność

Z drugiej strony, rzetelność to zdolność do odtworzenia spójnego wyniku w czasie i przestrzeni lub od różnych obserwatorów (Rysunek 6), przedstawiająca aspekty spójności, stabilności, równoważności i jednorodności. Jest to jedno z kryteriów jakości instrumentu (de Souza et al. 2017). Rzetelność dotyczy również powtarzalności. Na przykład o skali lub teście mówi się, że jest rzetelny, jeśli powtórzony pomiar dokonany za jego pomocą w stałych warunkach da ten sam wynik (Taherdoost H. 2016).

Należy podkreślić, że rzetelność nie jest stałą właściwością kwestionariusza. Wręcz przeciwnie, rzetelność zależy od funkcji instrumentu, populacji, w której jest używany, od okoliczności, od kontekstu; to znaczy, że ten sam instrument może nie być uznany za rzetelny w różnych warunkach. Na szacunkową rzetelność wpływa kilka aspektów środowiska oceny (osoby oceniające, charakterystyka próby, rodzaj instrumentu, metoda podawania) oraz zastosowana metoda statystyczna. Dlatego też wyniki badań z wykorzystaniem instrumentów pomiarowych mogą być interpretowane tylko wtedy, gdy warunki oceny i podejście statystyczne są jasno przedstawione. Podsumowując, rzetelność odnosi się do tego, czy instrument oceny daje takie same wyniki za każdym razem, gdy jest używany w tym samym otoczeniu z tym samym typem badanych (Sullivan G.M. 2011).

Wybór testów statystycznych wykorzystywanych do oceny rzetelności może być różny, w zależności od tego, co ma być mierzone. Istnieją trzy ważne kryteria rzetelności, które cieszą się dużym zainteresowaniem badaczy: 1) stabilność, 2) spójność wewnętrzną i 2) ekwiwalencja.



Rysunek 6 - W celu po lewej stronie strzały trafiły nie wiarygodnie, ponieważ trafione punkty nie są zlokalizowane w konkretnym miejscu, ale były rozrzucone po całym celu. W celu po prawej stronie strzały były niezawodne, trafiając w ten sam punkt (Image from De Souza A. et al. 2017).

Jaka jest procedura pomiaru wiarygodności instrumentu?

Jak wspomniano powyżej, statystyczna procedura pomiaru wiarygodności zależy od tego, co ma być mierzone.

1) **Stabilność:** stabilność mierzy, jak podobne są wyniki przy pomiarze w dwóch różnych momentach, czyli szacuje spójność powtarzalności pomiaru. Ocena stabilności może być przeprowadzona metodą test-retest, tzn. procedura polega na zastosowaniu tego samego pomiaru w dwóch różnych momentach. Współczynnik korelacji wewnątrzklasowej (ICC) jest jednym z najczęściej stosowanych testów do oceny stabilności zmiennych ciągłych, ponieważ uwzględnia błędy pomiarowe. Dla interpretacji wyników, minimalne wartości 0,70 są uważane za zadowalające (de Souza et al. 2017).

2) **Spójność wewnętrzna:** spójność wewnętrzna (lub homogeniczność) pokazuje, czy wszystkie podczęści instrumentu mierzą tę samą cechę. Jest to ważna właściwość pomiaru dla instrumentów oceniających jeden konstrukt, które wykorzystują różnorodne pozycje. Niska spójność wewnętrzna może wskazywać, że pozycje mierzą różne konstrukty (de Souza et al. 2017). W analizie spójności wewnętrznej oblicza się różnicę między odpowiedziami pozycji z danego konstrukt. Test Cronbacha alfa jest często wykorzystywany do obliczania wartości korelacji między odpowiedziami narzędzia oceny. Wysoki szacunek rzetelności powinien być tak bliski 1, jak to tylko możliwe (Sullivan G.M. 2011).

3) **Równoważność:** równoważność jest stopniem zgodności dwóch lub więcej obserwatorów w odniesieniu do wyniku narzędzia. Najbardziej powszechnym sposobem oceny równoważności jest rzetelność międzyobserwacyjna, która wymaga niezależnego udziału dwóch lub więcej oceniających. Współczynnik Kappa jest miarą używaną do oceny rzetelności międzyobserwacyjnej, stosowaną do zmiennych kategoryalnych i ma maksymalną wartość 1. Im wyższa jest wartość współczynnika Kappa, tym wyższa będzie zgodność między oceniającymi (Sullivan G.M. 2011).

Przykład 1: Niezawodność stabilności

Technika instrumentalna

Badanie międzyosobnicze test-retest (różni oceniający)

Meng L. i wsp. (2020) badali, czy mniejsza zależność od ręcznej identyfikacji anatomicznych punktów orientacyjnych może poprawić między-ocenową wiarygodność kinematyki stawów w porównaniu z trzema kinematycznymi modelami chodu. Ten cel badania został podniesiony, ponieważ główne źródło błędów w wiarygodności analizy

chodu wynika z palpacji anatomicznych punktów orientacyjnych. Hipotezą badania było stwierdzenie, że funkcjonalny model klastyczny Strathclyde (SFCM) uzyska większą wiarygodność między ocenami niż modele anatomiczne: Plug-in gait (PIG) oraz Human body model gait (HBM2). Aby to wykazać, 10 uczestników ukończyło trzy próby prowadzone przez różnych oceniających w nie następujących po sobie dniach. W każdej sesji, oceniający nakładał na uczestników zestaw połączonych markerów. Następnie, przez 5 s rejestrowano próbę statyczną, w której uczestnik stał w naturalnej postawie. Następnie każdy z uczestników szedł na bieżni z komfortową dla siebie prędkością. Po dwuminutowym okresie zapoznawczym rejestrowano trzydziestosekundowe dane, a zaznaczone trajektorie były rejestrowane przez 12 kamer z częstotliwością próbkowania 100 Hz. Współczynnik korelacji wewnątrzklasowej w cyklach chodu został użyty do porównania wiarygodności międzyaspektowej, a wartość ICC (α) została zinterpretowana na cztery poziomy wiarygodności: $\alpha < 0.5$ = słaba, $0.5 < \alpha < 0.75$ = umiarkowana, $0.75 < \alpha < 0.9$ = dobra, i $\alpha > 0.9$ = doskonała. Wyniki pokazują, że wszystkie modele wykazały "dobrą do doskonałej" wiarygodność między-aspektralną dla wszystkich kątów zgięcia/wyroztu i kąta odwodzenia/ przywodzenia biodra, ale wykazały "słabą do umiarkowanej" wiarygodność między-aspektralną dla innych kątów nie strzałkowych. Jednakże, SFCM uzyskał wyższą wiarygodność przy mniejszej zmienności w porównaniu z modelami anatomicznymi. Wyniki pokazują, że SFCM może być bardziej korzystny dla mniej doświadczonych oceniających.



Rysunek 7 - Układ markerów w modelu strathclyde functional cluster (po lewej) oraz połączony układ markerów dla modelu SFCM, PiG i HBM2 (po prawej).

Test-retest wewnątrz jednostki (różne czasy oceny)

Geerse D. et al. (2020) określili wiarygodność testu-retestu dla HoloLens Microsoftu, zestawu słuchawkowego typu mixed-reality, który oprócz hologramów zapewnia bogate dane o położeniu głowy, które mogą być wykorzystane do oceny ilościowej różnych zadań, takich jak chodzenie. Aby zbadać, jakie są granice zgodności pochodnych parametrów chodu spatio-temporalnego w czasie powtarzania, 23 zdrowych młodych dorosłych, że chodzenie w wolnym, komfortowym i szybkim tempie został sklasyfikowany, jak również 24 osób z chorobą Parkinsona, że chodzenie w wybranej przez siebie prędkości. HoloLens

zapewnia dane pozycji i orientacji głowy w środowisku w trzech kierunkach, a dane pozycji zostały wykorzystane do wyprowadzenia parametrów chodu spatiotemporalnego. Wiarygodność test-retest oceniono za pomocą współczynnika korelacji wewnątrzklasowej dla zgodności bezwzględnej. Współczynniki ICC powyżej 0,60 i 0,75 oznaczały odpowiednio dobrą i doskonałą zgodność. Współczynniki ICC wskazywały na doskonałą wiarygodność obu systemów dla zdrowych, młodych osób dorosłych oraz osób z chorobą Parkinsona.

			Trial 1		Trial 2	
			Mean ± SD	Mean ± SD	ICC _(A,1)	
Walking speed (cm/s)	HYA	SWS	74.2 ± 12.6	76.2 ± 12.2	0.861	
		CWS	116.2 ± 16.2	122.1 ± 16.4	0.870	
	PD	FWS	176.3 ± 21.8	174.7 ± 22.8	0.930	
		CWS	104.5 ± 20.7	106.1 ± 20.8	0.935	
Step length (cm)	HYA	SWS	56.2 ± 6.4	56.8 ± 5.4	0.884	
		CWS	67.9 ± 7.0	69.7 ± 6.6	0.911	
	PD	FWS	86.2 ± 8.0	85.5 ± 8.3	0.928	
		CWS	57.8 ± 11.5	57.8 ± 10.9	0.939	
Cadence (steps/min)	HYA	SWS	79.2 ± 9.8	80.6 ± 9.6	0.903	
		CWS	103.8 ± 8.5	106.5 ± 8.7	0.890	
	PD	FWS	124.1 ± 10.4	123.7 ± 9.8	0.953	
		CWS	109.6 ± 7.6	110.5 ± 7.0	0.778	

Ryc. 8 - Wiarygodność test-retest dla parametrów spatiotemporalnych chodu w warunkach instruowanego wolnego tempa chodu (SWS), komfortowego tempa chodu (CWS) i szybkiego tempa chodu (FWS) u zdrowych młodych dorosłych (HYA) i CWS u osób z chorobą Parkinsona (PD) uzyskanych z danych HoloLens (Geerse D. et al. 2020).

Zgodnie z oczekiwaniami, w analizach wiarygodności technik instrumentalnych, wartości ICC są wysokie, ponieważ techniki pomiarowe są zazwyczaj bardzo precyzyjne. Czynniki takie jak przestrzeganie ścisłego protokołu oceny pomagają w utrzymaniu wysokiej wiarygodności testu-retestu. Jeśli titerator stosuje instrumentalną technikę pomiarową bez przestrzegania tego samego protokołu oceny w sposób znormalizowany dla różnych czasów oceny, jest prawdopodobne, że wartości ICC nie były takie, jak zaobserwowano w przykładowym badaniu. Można również zaobserwować, że wiarygodność techniki pomiarowej spada w kadencji uczestników z chorobą Parkinsona. Dzieje się tak, ponieważ ta grupa pacjentów zmienia kadencję w zależności od długości wykonywanego kroku, która jest zmieniana przez chorobę. Przykład ten jest przydatny do zwrócenia uwagi na to, że wiarygodność pomiaru, mimo że jest to technika instrumentalna, jest różna w różnych populacjach.

Testy kliniczne

Analiza rzetelności jest jeszcze bardziej ceniona przez personel kliniczny w powszechnie stosowanych testach klinicznych. Hee-jae Kim i wsp. (2016) zbadali rzetelność (Intra-rater

test-retest) prędkości chodu mierzonej na różnych dystansach i w różnych tempach u starszych Koreańczyków. Pięćdziesiąt cztery uczestniczki w wieku ≥ 70 lat zostały zrekrutowane z lokalnej społeczności emerytów. Szybkość chodu oceniano na 4, 6 i 10 metrach, a także w trybie zwykłego i szybkiego chodu. Uczestnicy zostali poinstruowani, aby szli ze startu stojącego w tempie, które było dla nich normalne i wygodne lub aby szli tak szybko, jak tylko mogą, aż dotrą do końca wyznaczonej ścieżki. Przeszkolony tester siedł za uczestnikiem i zatrzymywał pomiar czasu, gdy stopa uczestnika dotykała podłogi na końcu trasy chodu. Oznacza to, że prędkość chodu została obliczona na podstawie przebytego dystansu i czasu, jaki zajęło uczestnikom pokonanie go. Wyniki przedstawiono na rycinie 9.

	4 M	Normal pace 6 M	10 M
ICC (95% CI)	.715**	.861**	.902**
	4 M	Maximal pace 6 M	10 M
	.837**	.905**	.933*

Rycina 9 - Wiarygodność testu chodu (ICC) dla różnych dystansów i tempa.

Wyższe wartości ICC zaobserwowano przy najdłuższym dystansie chodu wynoszącym 10 metrów w porównaniu do 4 i 6 metrów. Ponadto, wartości ICC testu chodu przy maksymalnej prędkości były wyższe niż przy normalnym tempie. Autorzy zaznaczyli, że chociaż test chodu w tempie maksymalnym na dłuższym dystansie ma lepszą wiarygodność u osób starszych, to odległość testu i tempo należy rozpatrywać w zależności od celu pomiaru i klinicznego stanu zdrowia uczestników, a nie według kryterium wysokiego poziomu wiarygodności.

Przykład 2: Wiarygodność wewnętrznej spójności

Badanie kliniczne

W przeciwieństwie do narzędzi instrumentalnych, za pomocą których mierzymy konkretną zmienną, taką jak prędkość chodu, długość kroku czy kąt ruchu w stawie, testy kliniczne pozwalają nam ocenić chód z bardziej globalnego punktu widzenia. Tak jest w przypadku Functional Gait Assessment (FGA), miary zdolności równowagi podczas chodu, opracowanej w celu wyeliminowania efektu pułapu obserwowanego w Dynamic Gait Index (DGI). Z oryginalnych 8 zadań DGI dodano trzy przypuszczalnie trudniejsze zadania i usunięto 1 łatwiejsze (Beninato M. i Ludlow L., 2016). Pozycje w FGA to: 1) Chód na równej powierzchni, 2) Zmiana szybkości chodu, 3) Chód z horyzontalnymi zwrotami głowy, 4) Chód z pionowymi zwrotami głowy, 5) Chód i obrót pivot, 6) Krok nad przeszkodą, 7) Chód z wąską podstawą podparcia, 8) Chód z zamkniętymi oczami, 9) Ambulowanie do tyłu oraz 10) Kroki. W każdej pozycji oceniający może przyznać ocenę od 0 do 3, uzyskując łącznie maksymalnie 30 punktów.

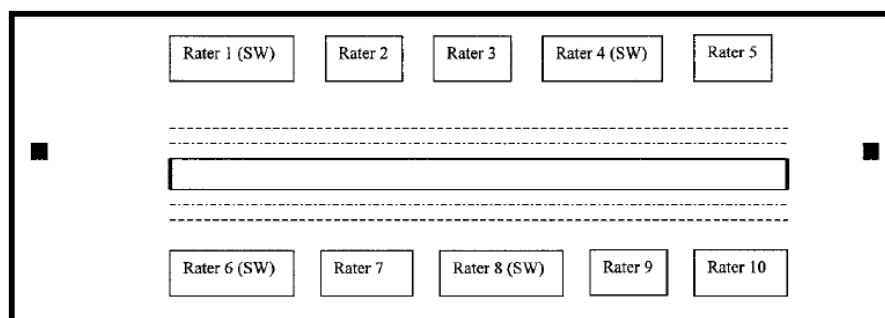
Wrisley D. i wsp. (2004), oceniali spójność wewnętrzną danych uzyskanych za pomocą FGA, gdy był on stosowany u osób z zaburzeniami przedsionkowymi. Sześciu pacjentów z zaburzeniami przedsionkowymi wypełniło FGA dwukrotnie, z godzinnym odpoczynkiem pomiędzy sesjami. Spójność wewnętrzną, czyli homogeniczność, pozycji zawartych w FGA określono za pomocą alfa Cronbacha. Ocenę tę przeprowadzono dla obu sesji testowych oraz w obrębie każdego z testów. Wartość alfa Cronbacha waha się od 0 do 1, spójność jest wysoka, gdy wartość alfa jest większa niż 0,8.

FGA wykazał wewnętrzną spójność w obrębie i pomiędzy obydwooma próbami testu FGA dla każdego pacjenta. Wartości alfa Cronbacha wynosiły odpowiednio .81 i .77 dla poszczególnych prób 1 i 2. Wartość alfa Cronbacha wynosiła .79 dla obu prób. Jest to jedna z analiz przeprowadzanych w celu walidacji narzędzia pomiarowego opartego na pozycjach. Zaletą tego typu testu jest to, że otrzymujemy wynik całościowej oceny funkcjonowania pacjentów w danym obszarze. W ten sposób uzyskujemy znacznie bardziej funkcjonalną miarę niż pojedynczy wynik mierzony techniką instrumentalną.

Przykład 3: Wiarygodność równoważności

Badanie kliniczne

W tym samym badaniu Wrisley D. i wsp. (2004) oceniali również równoważność lub stopień zgodności dwóch lub więcej obserwatorów w odniesieniu do Funkcjonalnej Oceny Chodu (FGA) za pomocą współczynnika Kappa. Do udziału w badaniu zgłosiło się siedmiu terapeutów fizycznych z różnych ośrodków oraz trzech studentów fizykoterapii. Ustawienie dla tej oceny było takie, jak pokazano na rycinie 10, gdzie wszyscy oceniający obserwowali tego samego pacjenta.



Ryc. 10 - Ustawienie chodnika i oznaczeń chodnika, plus pozycja oceniających do podawania testu. Czarne kwadraty reprezentują stożki wskazujące punkty startowe i stropowe. Ratownicy SW używali stropwatch podczas testu.

Statystyka kappa została użyta do oceny zgodności międzyosobniczej (między osobami oceniającymi) dla poszczególnych pozycji FGA i całości FGA. Wykres 11 zawiera procenty zgodności i wartości Kappa dla każdej pozycji i całkowitych wyników FGA. Cohen zasugerował, aby wynik Kappa był interpretowany w następujący sposób: wartości ≤ 0 jako brak zgodności, 0,01-0,20 jako brak lub niewielka zgodność, 0,21-0,40 jako rzetelna, 0,41-0,60 jako umiarkowana, 0,61-0,80 jako znaczna, a 0,81-1,00 jako prawie doskonała

zgodność (McHugh M. 2012).

	FGA Item										Total
	1	2	3	4	5	6	7	8	9	10	
Interrater reliability											
% agreement	88	60	58	68	60	68	66	66	67	90	58
Kappa	.78	.37	.40	.53	.34	.41	.45	.46	.54	.76	.50

Ryc. 11 - Wiarygodność Interrateralna dla poszczególnych i całosciowych elementów Funkcjonalnej Oceny Chodu (FGA).

Zgodność została określona pomiędzy 10 klinicystami, którzy zmierzili 6 pacjentów podczas pierwszej próby pomiarowej. Ponieważ istnieje 45 możliwych porozumień między parami klinicystów na jednego uczestnika, u 6 uczestników uzyskano 270 możliwych porozumień dla każdej pozycji FGA i całkowitej punktacji FGA. Wartości na rycinie 11 przedstawiają średnią zgodność i kappa dla 45 par klinicystów dla każdej pozycji. Średnia zgodność test-retest pomiędzy wszystkimi parami klinicystów, w naszej opinii, była umiarkowana lub lepsza dla wszystkich pozycji z wyjątkiem pozycji 2 ("zmiana prędkości chodu") i 5 ("chód i obrót obrotowy").

Wyniki dotyczące rzetelności międzylaboratoryjnej w badaniu Wrisley D. i wsp. (2004) są użyteczne jako przykład tego, że skale i testy, które są opracowywane na podstawie obserwacji osoby oceniającej, mogą różnić się w wynikach w zależności od tego, co osoba oceniająca obserwuje, co jest uważane za wadę klinicznych skal i testów w porównaniu z narzędziami instrumentalnymi.

2.6 Wrażliwość na zmiany i reaktywność

Wrażliwość na zmianę definiuje się jako zdolność instrumentu do pomiaru zmiany stanu, niezależnie od tego, czy zmiana ta jest istotna lub znacząca dla decydenta. Czułość na zmianę, choć konieczna, została opisana jako niewystarczająca do oceny zmiany i ustalenia skuteczności leczenia. Test może być wrażliwy na stan lub diagnozę, ale to, czy jest on znaczący lub ważny, nie może być wywnioskowane tylko z tej właściwości.

Znaczenie tej właściwości przyrządów do pomiaru chodu jest związane z oceną wpływu programów i terapii w naukach klinicznych. Jeśli instrument nie jest wystarczająco czuły, nie będzie w stanie ocenić wpływu określonej patologii lub leczenia na chód człowieka. Taka czułość jest szczególnie istotna w warunkach stosowanych, gdzie efekty programu lub leczenia często nie są szczególnie silne, a warunki pomiaru mogą być dość zmienne (Lipsey

M. 1983). Jeśli instrument nie jest wystarczająco czuły, aby wykryć zmianę (wielkość efektu) w pewnych warunkach lub leczeniu, często potrzebne są duże próby, aby osiągnąć odpowiednią moc statystyczną w celu wykrycia efektu tych interwencji.

Z drugiej strony, responsywność jest definiowana jako zdolność instrumentu do pomiaru istotnej lub klinicznie ważnej zmiany stanu klinicznego i jest uznawana za podstawową właściwość instrumentów przeznaczonych do pomiaru zmiany i skuteczności interwencji. Podobnie jak rzetelność i trafność, responsywność nie jest uważana za cechę, którą można uogólnić i powinna być oceniana dla każdej populacji i celu, dla którego dany instrument jest używany.

Istotna klinicznie lub ważna zmiana może być zdefiniowana, a zatem oceniana z perspektywy pacjenta, jego pełnomocnika, społeczeństwa lub pracownika służby zdrowia. Oznacza to zmianę, która jest zauważalnie, odczuwalnie różna, która ma wartość dla pacjenta (lub lekarza). Zmiana ta może pozwolić danej osobie na wykonanie jakiegoś istotnego zadania lub na wykonywanie zadań w sposób bardziej efektywny, z mniejszym bólem lub trudnością. Zmiany te powinny również przekraczać zmienność, którą można przypisać przypadkowi. Jaeschke i wsp. (1989) zasugerowali, że klinicznie istotna zmiana może być zdefiniowana jako minimalna istotna różnica. Może to być zdefiniowane jako najmniejsza różnica w punktacji w domenie zainteresowania, którą pacjent postrzega jako zmianę i która nakazałaby, przy braku efektów ubocznych i nadmiernych kosztów, modyfikację postępowania z pacjentem.

W ten sposób responsywność jest powszechnie raportowana poprzez oszacowanie minimalnie istotnej różnicy (MID) lub minimalnej klinicznie istotnej różnicy (MCID), przy czym zmiana wyniku w pomiarze powinna być równa lub przekraczać oszacowanie MID, aby mogła być uznana za istotną. Oszacowanie MID miar ułatwia interpretację wyniku zmiany, ustanawiając punkty odniesienia, które pomagają określić istotność zmiany.

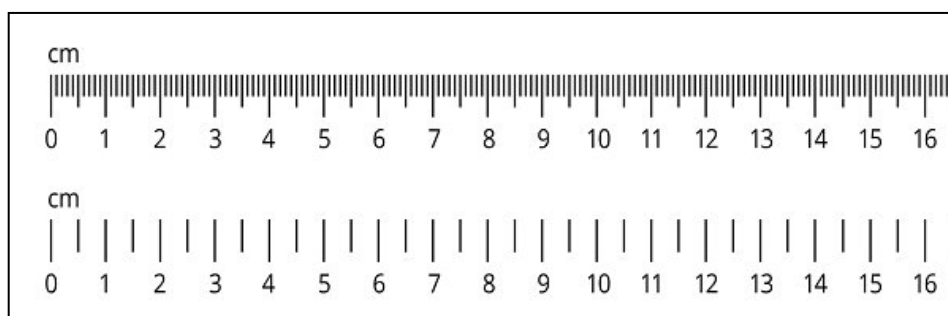
Podjęcie praktyczne

Oceniając przydatność kliniczną terapii mających na celu poprawę subiektywnych wyników leczenia, należy określić wielkość poprawy, która jest istotna dla pacjentów. Minimalna istotna klinicznie różnica (minimal clinically important difference - MCID) to najmniejsza korzyść, która ma wartość dla pacjenta. MCID jest pojęciem skoncentrowanym na pacjencie, obejmującym zarówno wielkość poprawy, jak i wartość, jaką pacjenci przypisują tej zmianie. Innymi słowy, MCID definiuje najmniejszą kwotę, o jaką musi zmienić się wynik, aby był znaczący dla pacjentów (McGlothlin A. i wsp. 2014).

W 2014 roku Bohannon R. i wsp. opracowali przegląd systematyczny na temat Minimal clinically important difference for change in comfortable gait speed of adults with pathology. Wniosek z badania był taki, że zmiany w prędkości chodu rzędu 0,10 do 0,20 ms⁻¹ mogą być istotne w wielu grupach pacjentów, takich jak osoby m.in. z udarem mózgu, złamaniem biodra, stwardnieniem rozsianym. Dlatego też, jeśli chcemy zmierzyć jakąkolwiek interwencję lub ewolucję u osób z patologiami wymienionymi w poprzednim badaniu, musimy używać sprzętu pomiarowego o wystarczającej czułości do rejestrowania zmian prędkości chodu, takich jak te wskazane w badaniu Bohannon R. i wsp., czyli zdolnego do rejestrowania zmian prędkości rzędu 0,1 ms⁻¹.

Podsumowując, pracownicy służby zdrowia, wybierając urządzenie pomiarowe, powinni brać pod uwagę, że czułość sprzętu musi być wystarczająca do pomiaru minimalnej klinicznie istotnej różnicy w zmiennej, którą chcą obserwować w danej populacji.

W odniesieniu do czułości, biomechaniczne techniki oceny mają większą czułość w wykrywaniu zmian w cechach chodu niż kliniczne skale oceny. Dzieje się tak dlatego, że biomechaniczne instrumenty pomiarowe są znacznie dokładniejszym sprzętem. Aby zrozumieć tę koncepcję, spójrz na Rycinę 12.



Rysunek 12 - Obraz przedstawia dwie linijki. Górna linijka jest wyskalowana w mm, a dolna w cm, a ich dokładności wynoszą odpowiednio ± 1 mm i ± 1 cm = ± 10 mm. Ponieważ pierwsza linijka jest dokładniejsza od drugiej, górna linijka jest bardziej czuła niż dolna.

Górna linijka jest dokładna do 1mm, a dolna linijka na rysunku jest dokładna do 1cm (10mm). Ponieważ górna linijka jest bardziej precyzyjna niż dolna, jest ona z kolei bardziej wrażliwa na pomiar długości. Tak samo jest w przypadku biomechanicznych przyrządów do pomiaru chodu człowieka. Na rysunku 5 przedstawiono precyzję różnych instrumentów biomechanicznych do pomiaru parametrów przestrzenno-czasowych podczas chodu człowieka. Systemy fotogrametryczne i kamery optoelektroniczne są najdokładniejszymi urządzeniami w porównaniu z innymi systemami, dlatego są one również bardziej czułe na pomiar zmiennych przestrzenno-czasowych niż chody wyposażone w czujniki ciśnienia, czujniki inertialne lub stoper (Moissenet F. et al.). Niestety, dowody naukowe dotyczące czułości urządzeń pomiarowych są ograniczone.

Przykład: czułość testu klinicznego chodu u osób z uszkodzeniem rdzenia kręgowego

Tak jak narzędzia oceny biomechanicznej różnią się wrażliwością na pomiar określonych wyników, skale kliniczne również posiadają tę cechę. Jackson A. i wsp. (2008) przeanalizowali przydatność w praktyce klinicznej obecnych miar wyników stosowanych jako wskaźniki poprawy chodu i poruszania się w populacji osób z uszkodzeniem rdzenia kręgowego. Konkretnie, oceniali następujące miary chodu i poruszania się: indeks chodu dla uszkodzeń rdzenia kręgowego II (WISCIII), test chodu na 50 stóp (50FTWT), test 6-minutowego chodu (6MWT), test 10-metrowego chodu (10MWT) oraz funkcjonalny pomiar niezależności - lokomocja (FIM-L). Wyniki przedstawiono w tabeli 4.

Tabela 4 - Wrażliwość na zmiany w sprawności lokomocyjnej osób z uszkodzeniem rdzenia kręgowego

Test kliniczny	Czułość
Wskaźnik chodu dla uszkodzeń rdzenia kręgowego II (WISCI II)	<ul style="list-style-type: none"> - Dobra wrażliwość na zmiany u pacjentów z bardziej upośledzonym chodem. - Nie obejmuje elementów szybkości lub wytrzymałości.
Test chodu na 50 stóp (50FTWT)	<ul style="list-style-type: none"> - Osoby, które nie są w stanie przejść 50 stóp, nie są w stanie uczestniczyć w pomiarze (efekt dna). - Pacjenci, którzy są w stanie przejść dystans w szybkim tempie, mogą również być w stanie przejść większy dystans w tym samym tempie (efekt pułapu).
Test 6-minutowego marszu (6MWT)	<ul style="list-style-type: none"> - Dobra wrażliwość na zmiany u osób z lepszą funkcją ambulatoryjną w ostrym urazie i 6 miesięcy po urazie (podczas gdy WISCI II i Lower Extremity Motor Score nie). - U pacjentów z udarem test wykazał mniejszą czułość na zmiany niż 12 MWT. - Na czułość testu 6MWT wpływa efekt dna wśród pacjentów, którzy nie mogą chodzić przez 6 minut. - Podczas testu dozwolone jest odpoczywanie w pozycji stojącej, ale jeśli badany siedzi, jest dyskwalifikowany. Podobnie, efekt pułapu obserwuje się u pacjentów, którzy mogą kontynuować chód powyżej 6 minut w tym samym tempie.
Test chodu na 10 metrów (10MWT)	<ul style="list-style-type: none"> - Większa czułość u pacjentów z ponad 90% poprawą wyniku motorycznego kończyn dolnych w 6 i 12 miesiącu po urazie (ASIA D). - Lepsza czułość niż WISCI II, szczególnie u osób z mniejszą niepełnosprawnością, ze względu na efekt pułapowy tego testu. - Czułość na zmiany w chodzie jest bardziej porównywalna z innymi testami szybkości chodu, takimi jak 6MWT, ponieważ mierzone są podobne cechy. - Wrażliwość na zmiany wykazano również w populacji osób po udarze mózgu, stosując zmodyfikowany dystans 6 m jako odmianę 10MWT. - Większa czułość niż w przypadku wskaźnika motoryczności kończyn dolnych w dłuższym okresie (6 miesięcy). - Miara wyników jest mniej wrażliwa na zmiany w lokomocji w przewlekłym SCI, gdzie zmiany w sile są niewielkie.
Pomiar niezależności funkcjonalnej lokomotorycznej (FIM-L)	<ul style="list-style-type: none"> - Słaba wrażliwość na wykazanie zmiany w wyniku jednego uczestnika przed i po treningu, pomimo zmiany sposobu poruszania się z wózka inwalidzkiego na chodzenie. - Efekt pułapowy ujawnia się w tym, że test wykazuje słabą wrażliwość na zmiany u osób z lepszymi umiejętnościami chodzenia. Ponadto, punktacja FIM-L nie uwzględnia szybkości, z jaką badani pokonują wymagane dystanse.

Wniosek z badania był taki, że u osób z uszkodzeniem rdzenia kręgowego test chodu na 10 metrów i Walking Index for Spinal Cord Injury II są najbardziej wiarygodnymi i klinicznie użytecznymi testami w ocenie poprawy chodu i ambulacji, między innymi ze względu na dobrą czułość w ocenie zmian w ambulacji w tej grupie pacjentów, odpowiednio z wysokim stopniem powrotu do zdrowia i z bardziej zmienionym chodem.

W poprzednim przykładzie widzieliśmy, że na czułość testów oceny chodu mają wpływ dwa zjawiska zwane efektem sufitowym i efektem podłogowym, które są związane z możliwością wykonania samego testu przez osoby, do których jest on skierowany.

2.7 Efekt podłogi i sufitu

Jeśli zakres funkcji objętych pomiarem jest mniejszy niż zakres odczuwany przez pacjentów, pomiarowi może brakować zdolności reagowania. Potencjał efektów dolnego i górnego pułapu jest często oceniany poprzez badanie wzorców odpowiedzi. Jeśli występują skoki przy najwyższej lub najniższej opcji odpowiedzi, jest to często interpretowane jako dowód na istnienie odpowiednio efektów pułapu lub dna. Jednakże, kiedy używa się miar do oceny skuteczności interwencji, prospektywne dowody na działanie miary są ważniejsze niż to, czy występują skoki (Feeny DH. et al. 2013).

Jeśli wyniki uczestników skupiają się w kierunku górnej granicy (lub najlepszego możliwego wyniku) środka/instrumentu (efekt pułapu) lub w dolnej granicy (efekt dna), zmiany doświadczone przez pacjentów w wydajności chodu mogą być tendencyjne. Wynika to z faktu, że pacjenci mogą być w "gorszej sytuacji" niż ta, którą może uchwycić miara lub w "lepszey sytuacji" niż ta, którą może zmierzyć instrument.

Jeśli porównamy instrumentalne techniki oceny z klinicznymi skalami oceny, zauważymy, że techniki instrumentalne mają efekt dolny, podczas gdy kliniczne skale oceny chodu mają efekt górny. Powodem tej różnicy jest łatwość lub trudność, z jaką każda z nich może być wykonana przez pacjenta. Skale oceny są definiowane na podstawie ustrukturyzowanych kwestionariuszy, dokonując klasyfikacji kategoriowych lub dyskretnych, które zmniejszają czułość w odniesieniu do innych instrumentów lub technologii oceny klinicznej. Innymi słowy, małe zmiany, które zachodzą w funkcjonalnych możliwościach pacjenta w wyniku interwencji specjalisty, są bardzo trudne do zidentyfikowania. Fakt ten implikuje, w zależności od konstrukcji skali, efekt sufitowy. Jednakże, osoby z poważnymi zaburzeniami chodu nie będą w stanie wykonać złożonych protokołów oceny wymaganych przez wiele biomechanicznych technik instrumentalnych, więc najbardziej czułą oceną dla tego typu pacjentów będą skale oceny klinicznej.

Z drugiej strony, instrumentalne techniki oceny biomechanicznej mają efekt dna, ponieważ protokoły oceny wymagają większego zapotrzebowania ze strony pacjenta do ich wykonania. Oznacza to, że w większości przypadków ocena chodu osób z poważnym upośledzeniem zdolności poruszania się nie jest możliwa przy użyciu technik instrumentalnych, dlatego też

efekt dna może ograniczyć wejście na rynek osób z poważnymi urazami do czasu, gdy będą one w stanie chodzić z wymaganym oprzyrządowaniem.

Przykład

Chociaż badania, które analizują właściwości statystyczne instrumentów wartościujących są ograniczone, Middleton A. i Fritz S. (2013), przedstawili dowody dotyczące dokładnej oceny chodu u starszych dorosłych, ponieważ jest to ważny aspekt praktyki klinicznej dla klinicystów pracujących z tą populacją. Wybrane miary są następujące: Chód: szybkość chodu, symetria chodu, wytrzymałość chodu, zdolność adaptacyjna chodu, wydajność podwójnego zadania podczas chodu oraz samoocena pewności siebie podczas chodu. Rycina 13 przedstawia podsumowanie miar wyników uwzględnionych w artykule jako szybki punkt odniesienia dla klinicystów.

Measure	Assessing	Predictive abilities	Scale ^a	MDC ₉₅
4-Meter Walk Test [20•, 21, 24, 25, 33••, 35]	Gait speed	Functional status, discharge location, rehabilitation potential, fall risk, mortality	n/a	0.14 m/s
6-Minute Walk Test [35, 38]	Gait endurance	Not established	n/a	58.2 m
Dynamic Gait Index [39]	Ability to adapt to changing task demands during gait	Fall risk	0-24	2.9 points
Walking While Talking Test [48]	Gait performance under divided attention conditions	Fall risk, frailty, disability, mortality	n/a	Not established
Modified Gait Efficacy Scale [49•]	Patients' perceptions of gait abilities	Not established	10-100	14.7 points

Ryc. 13 - Tabela zbiorcza miar wyników chodu. Minimalna wykrywalna zmiana (MDC), nie dotyczy (n/d).

Minimalna wykrywalna zmiana (MDC) zapewnia klinicystom wartość referencyjną do określenia, czy wystąpiła prawdziwa zmiana. MDC₉₅ określa ilościowo najmniejszą wielkość zmiany wymaganą do przekroczenia błędu pomiaru i zmienności przy poziomie ufności 95%. Wyniki MDC dla miar wyników, jeśli są opisane w literaturze, mogą być cennym narzędziem dla klinicystów. Na rycinie 13 można zaobserwować minimalną wykrywalną zmianę dla wyników chodu najczęściej stosowanych w praktyce klinicznej, mierzonych za pomocą instrumentów wymienionych w pierwszej kolumnie.

Badanie to jest przydatne, aby wyjaśnić, że pacjenci, którzy osiągają wyniki poniżej minimalnej wykrywalnej zmiany w każdej z wymienionych zmiennych, preferowane jest, aby byli mierzni za pomocą innych instrumentów, które pozwalają na wykrycie zmian poniżej wartości referencyjnej jako minimalnej wykrywalnej zmiany.

3. Kluczowe idee

- Personel medyczny musi znać charakterystykę metodologiczną i właściwości statystyczne w momencie wyboru narzędzia do oceny chodu. Jest to konieczne, aby uniknąć błędów metodologicznych i tendencyjności wyników pomiaru.
- Jeśli chodzi o użyteczność, skale i testy kliniczne mają tę zaletę, że można je opracować w krótkim czasie, nie wymagają specjalistycznego szkolenia od oceniającego i mogą być stosowane w każdym kontekście, np. w praktyce klinicznej.
- Sprzęt potrzebny do użycia testów klinicznych i skal jest znacznie mniejszy i łatwiej dostępny niż sprzęt potrzebny do przeprowadzenia oceny chodu za pomocą przyrządów do oceny biomechanicznej.
- Najważniejszą cechą instrumentalnych technik oceny biomechanicznej jest to, że dostarczają one obiektywnych danych uzyskanych bez interpretacji oceniającego (tj. bezpośrednia ocena jednego lub więcej wymiarów wzorca chodu), więc ich zastosowanie jest głównie w obszarze badawczym. Przeciwnie, informacje uzyskane za pomocą skal i testów klinicznych są pod wpływem interpretacji i percepcji osoby oceniającej.
- Wysoka precyzja instrumentalnych technik pomiarowych daje im cechę bycia bardziej wiarygodnymi w pomiarze cech chodu niż skale i testy kliniczne.
- Wiarygodność jest zazwyczaj lepsza w przypadku przyrządów biomechanicznych, ponieważ powtarzalność pomiaru nie zależy od obserwatora, ale od innych czynników, takich jak wykonanie pomiaru z zastosowaniem standaryzowanego protokołu.
- Im dokładniejszy jest przyrząd pomiarowy, tym bardziej wrażliwy na zmiany będzie ten przyrząd. Czułość sprzętu musi być wystarczająca do zmierzenia minimalnej klinicznie istotnej różnicy w wynikach, które profesjonalista zamierza zaobserwować w danej populacji.
- Skale kliniczne i testy mają większą tendencję do występowania efektu pułapu, to znaczy, że wyniki uczestników skupiają się w kierunku wysokiego końca (lub najlepszego możliwego wyniku) środka / instrumentu. Z drugiej strony, techniki instrumentalne mają większy efekt dolnej granicy, gdzie wyniki uczestników skupiają się w kierunku dolnej granicy. Jest to spowodowane tym, że pacjenci mogą być w "lepszej sytuacji" niż ta, którą może uchwycić pomiar lub w "gorszej sytuacji" niż ta, którą może zmierzyć instrument.

4. Bibliografía

- [1] Gutierrez-Clavería M, Beroíza T, Cartagena C, Caviedes I, Céspedes J, Gutiérrez-Navas M, Oyarzún M, Palacios S, Schönfeldt P. Guidelines for the six-minute walk test. *Rev Chil Enf Respir* 2009; 25: 15-24.
- [2] Innerd P, Catt M, Collerton J, Davies K, Trenell M, Kirkwood T, Jagger C. A comparison of subjective and objective measures of physical activity from the Newcastle 85+ study. *Age Ageing*. 2015 Jul;44(4):691-4.
- [3] Crémers J, Phan R, Delvaux V, Garrauxa G. Construction and validation of the Dynamic Parkinson Gait Scale (DYPAGS). *Parkinsonism & Related Disorders*. Volume 18, Issue 6, July 2012, Pages 759-764.
- [4] Tinetti M.E. Performance-Oriented Assessment of Mobility Problems in Elderly Patients. *J Am Geriatr Soc*. 1986 Feb;34(2):119-26.
- [5] Wrisley D, Kumar N. Functional Gait Assessment: Concurrent, Discriminative, and Predictive Validity in Community-Dwelling Older Adults. *Phys Ther*. 2010 May;90(5):761-73.
- [6] Pinto R, Birmingham T, Leitch K, Atkinson H, Jones I, Giffin J.R. Reliability and validity of knee angles and moments in patients with osteoarthritis using a treadmill-based gait analysis system. *Gait & Posture* 80 (2020) 155-161.
- [7] Taherdoost H. Validity and Reliability of the Research Instrument: How to Test the Validation of a Questionnaire/Survey in a Research. *International Journal of Academic Research in Management (IJARM)*. Vol. 5, No. 3, 2016, Page: 28-36.
- [8] De Souza A, Costa N, de Brito E. Psychometric properties in instruments evaluation of reliability and validity. *Epidemiol. Serv. Saude, Brasília*, 26(3), Jul-Sep 2017.
- [9] Sullivan G. A primer on the Validity of Assessment Instruments. *J Grad Med Educ*. 2011.
- [10] Meng L, Millar L, Childs C, Buis A. A strathclyde cluster model for gait kinematic measurement using functional methods: a study of inter-assessor reliability analysis with comparison to anatomical models. *Computer methods in biomechanics and biomedical engineering*. *Comput Methods Biomech Biomed Engin*. 2020 Jun 16;1-10.
- [11] Geerse D, Coolen B, Roerdink M. Quantifying Spatiotemporal Gait Parameters with HoloLens in Healthy Adults and People with Parkinson's Disease: Test-Retest Reliability, Concurrent Validity, and Face Validity. *Sensors (Basel)*. 2020 Jun 5;20(11):3216.
- [12] Hee-jae Kim, Ilhyoek Park, Hyo joo Lee, On Lee. The reliability and validity of gait speed with different walking pace and distances against general health, physical function, and chronic disease in aged adults. *J Exerc Nutrition Biochem*. 2016;20(3):046-050.

- [13] Wrisley D, Marchetti G, Kuharsky D, Whitney S. Reliability, Internal Consistency, and Validity of Data Obtained With the Functional Gait Assessment. *Phys Ther.* 2004 Oct;84(10):906-18.
- [14] McHugh M. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012 Oct; 22(3): 276-282.
- [15] Lipsey, M. W. (1983). A scheme for assessing measurement sensitivity in program evaluation and other applied research. *Psychological Bulletin*, 94(1), 152–165.
- [16] Jaeschke R, Singer J, Guyatt GH. Ascertaining the minimal clinically important difference. *Cont Clin Trials.* 1989;10:407–415.
- [17] McGlothlin A. and Lewis R. Minimal Clinically Important Difference Defining What Really Matters to Patients. *JAMA* October 1, 2014 Volume 312, Number 13.
- [18] Bohannon R and Glenney S. Minimal clinically important difference for change in comfortable gait speed of adults with pathology: a systematic review. *Journal of Evaluation in Clinical Practice* 20 (2014) 295–300.
- [19] Moissenet F, Armand S. Chapter 17: Qualitative and quantitative methods of assessing gait disorders. *Orthopedic Management of Children with Cerebral Palsy.* 2015 Ed. Nova Science Publishers, Inc. ISBN: 978-1-63483-318-9-
- [20] Jackson A, Carnel C, Ditunno J, Schmidt M. Boninger M, Schmeler M, Williams S, Donovan W. Outcome Measures for Gait and Ambulation in the Spinal Cord Injury Population. *J Spinal Cord Med.* 2008;31:487–499.
- [21] Feeny DH, Eckstrom E, Whitlock EP, Perdue LA. Agency for Healthcare Research and Quality, US. A Primer for Systematic Reviewers on the Measurement of Functional Status and Health-Related Quality of Life in Older Adults. September 2013.
- [22] Middleton A, Fritz S. Assessment of Gait, Balance, and Mobility in Older Adults: Considerations for Clinicians. *Curr Transl Geriatr and Exp Gerontol Rep* (2013) 2:205–214.



Wsparcie Komisji Europejskiej dla produkcji tej publikacji nie stanowi poparcia dla treści, które odzwierciedlają jedynie poglądy autorów, a Komisja nie może zostać pociągnięta do odpowiedzialności za jakiegokolwiek wykorzystanie informacji w niej zawartych.

