

Piotr Brzyski, Beata Tobiasz-Adamczyk, Tomasz Knurowski
Zakład Socjologii Medycyny, Katedra Epidemiologii i Medycyny Zapobiegawczej,
Uniwersytet Jagielloński, Collegium Medicum w Krakowie

Trafność i rzetelność skali GARS w populacji osób w starszym wieku w Polsce

*Validity and reliability of the GARS scale in population
of older people in Poland*

Abstract

Introduction. Measurement scale, which was created outside Poland, should not be used in sample of Polish older people without previous assessment of its validity and reliability. It is necessary due to possible differences in meaning or understanding the items included in the scale measuring health-related quality of life. Such procedure is also necessary in case of scale measuring functional status, though it seems that wording of the items does not contain words and phrases which are strongly influenced by elements specific for the culture of the country in which the scale was developed.

Material and methods. The study was conducted in the simple random sample of 528 community dwelling citizens of Krakow aged 65–85. Analysis of psychometric properties of the GARS in terms of classical test theory was conducted with principal component analysis (validity) and Cronbach alpha coefficient (internal consistency reliability). Analysis of psychometric properties of the scale in terms of item response theory was conducted with Mokken scaling procedure: reliability coefficient rho and scalability coefficient H Loewinger were assessed.

Results. All items of the GARS scale correlated with first principal component higher than 0,56, supporting thesis about unidimensional structure of the scale. Reliability of the GARS scale was high, both in terms of classical test theory ($\alpha = 0,92$) as well as in terms of item response theory ($\rho = 0,95$). Scalability coefficient H for the whole scale equaled 0,70, whereas for particular items ranged from 0,54 to 0,96.

Conclusions. The GARS scale is valid, highly reliable, and strongly homogenous tool for measuring functional status.
Gerontol. Pol. 2012; 20, 3: 109–117

Key words: validity, reliability, homogeneity, scalability, health-related quality of life, GARS, functional status, older people

Wstęp

Stan funkcjonalny, rozumiany jako zdolność samodzielnego wykonywania podstawowych czynności życiowych, odgrywa niezwykle istotną rolę w ogólnej ocenie jakości życia u osób w wieku starszym: decyduje

o takich aspektach życia, jak kontynuacja pełnionych ról społecznych, zakres interakcji społecznych czy też ogólne zadowolenie z życia. Stan funkcjonalny określa miejsce osoby starszej w różnych grupach społecznych i jest niezależnym predyktorem wpływającym na poziom istotnego wskaźnika jakości życia warunkowanej stanem zdrowia (HRQoL), jakim jest subiektywna ocena stanu zdrowia [1]. Badania nad predyktorami umieralności dowodzą, że stan funkcjonalny w starszym wieku jest niezależnym predyktorem ryzyka zgonu [2–5] oraz może być czynnikiem modyfikującym wpływ innych predyktorów,

Adres do korespondencji:
dr n. med. Piotr Brzyski
Zakład Socjologii Medycyny
Katedra Epidemiologii i Medycyny Zapobiegawczej UJ CM
ul. Kopernika 7A, 31–034 Kraków
tel.: 12 424 72 37, w. 15, faks: 12 422 87 95

takich jak np. wsparcie społeczne [6]. Stanowi on miernik stopnia niezależności osoby starszej wiekiem od innych osób i możliwości prowadzenia przez nią samodzielnego życia w środowisku zamieszkania. Wykonywanie bez ograniczeń czynności codziennego życia jest traktowane jako wskaźnik zdrowej starości i stawiane za cel wielu programów profilaktyki i promocji zdrowia [7].

Pomiar stanu funkcjonalnego, podobnie jak innych wymiarów stanu zdrowia oraz jakości życia, odbywa się najczęściej za pomocą skal. Są to narzędzia badawcze składające się z kilku lub kilkunastu pytań, wybranych na podstawie zaawansowanej metodologii i posiadających wystandaryzowany zestaw odpowiedzi — najczęściej w formacie Likerta. Większość skal badawczych, służących do pomiaru stanu zdrowia i jakości życia, opracowano poza Polską, najczęściej w krajach anglojęzycznych. Bezskrytyczne zastosowanie takiej skali pomiarowej w środowisku, w którym nie została ona utworzona, może prowadzić do zafałszowania wyników, co może mieć źródło w zależności wskaźników używanych do oceny tych samych zjawisk od warunków kulturowych lub być rezultatem różnic w znaczeniach tych samych symboli i wartości, wykorzystywanych jako wskaźniki konstruktywów teoretycznych mierzonych przez skalę w różnych kręgach kulturowych [8]. Konieczne zatem staje się zbadanie, czy wysokie walory narzędzia, spełniającego doskonale swoją rolę w warunkach, w których zostało stworzone, zostają zachowane przy jego adaptacji do nowych uwarunkowań kulturowych, oraz określenie, jakich zabiegów adaptacyjnych wymaga jego zastosowanie w nowym środowisku.

Aby określić właściwości psychometryczne skali mierzącej jakość życia warunkowaną stanem zdrowia w polskich warunkach kulturowych, należy zbadać w możliwie szerokim zakresie jej trafność i rzetelność. Trafność testu jest zagadnieniem złożonym, ponieważ wyróżnia się trzy jej aspekty: trafność kryterialną, treściową i teoretyczną [9]. Niniejsza praca dotyczy aspektu teoretycznego trafności, czyli związku narzędzia pomiarowego z konstruktem teoretycznym — zmienną ukrytą (latentną), którą skalą ma mierzyć [10]. Rzetelność jest miarą tego, w jakim stopniu wynik testu oddaje rzeczywistą wartość badanej cechy, a w jakim pomiar ten jest zakłócony przez błąd pochodzący z różnych źródeł [9, 11]. Według klasycznej teorii testu rzetelność jest definiowana jako udział wariancji wyniku prawdziwego zmiennej ukrytej mierzonej przez skalę w wariancji wyniku obserwowanego — wyniku pomiaru dokonanego za pomocą skali, gdzie jako wynik prawdziwy określa się

wynik, jaki uzyskałaby osoba badana testem idealnie mierzącym daną zmienną ukrytą w idealnych warunkach, eliminujących całkowicie błąd pomiaru [9, 12]. Istotnym aspektem oceny wartości psychometrycznej w przypadku skali GARS, podobnie jak w przypadku innych skal mierzących stan funkcjonalny, których wartość psychometryczną ocenia się także w odniesieniu do teorii odpowiedzi na pozycje testowe (IRT, *item response theory*), oprócz wspomnianych parametrów, jest jej homogeniczność (w literaturze anglojęzycznej określana jako *scalability*) [13, 14]. Homogeniczność oznacza, że skala mierzy ten sam czynnik wspólny u wszystkich badanych oraz we wszystkich pozycjach skali. Pod określeniem tym kryje się połączenie dwóch terminów: jednowymiarowości i hierarchiczności. Jednowymiarowość oznacza, że wszystkie pozycje skali mierzą ten sam konstrukt teoretyczny (zmienną ukrytą), hierarchiczność zaś oznacza istnienie zależności między pozycjami skali, która w idealnej postaci cechuje się tym, że pozwala na uporządkowanie pozycji skali według stopnia trudności oraz że osoby o tych samych wynikach odpowiadają w ten sam sposób na te same pytania, a jeśli odpowiedziały „poprawnie” na wybrane pytanie, to odpowiedziały „poprawnie” na pytania „łatwiejsze” od tego pytania [15, 16]. W przypadku skali oceniającej stan funkcjonalny jako poprawną odpowiedź rozumiemy zdolność do samodzielnego wykonania określonej czynności (wymagającej odpowiedniego wysiłku czy stopnia sprawności niezbędnego do jej wykonania), natomiast przez trudność pozycji rozumiemy ilość tego wysiłku, a kontekście interpretacji wyników skali mierzymy go średnią wyników uzyskanych na danej pozycji skali przez ankietowane osoby.

Metodologia badań nad wskaźnikami stanu funkcjonalnego wypracowała wiele różnorodnych skal mierzących sprawność funkcjonalną — do najbardziej znanych i zarazem najczęściej stosowanych należą: *Index of Independence in Activities of Daily Living* (ADL) [17] i *Instrumental Activity of Daily Living* (IADL) [18], a także *Barthel Indeks* [19], *OECD Disability Questionnaire* [20], *Functional Status Index* [21] oraz *MOS Physical Functioning Measure* [22]. Skale te zostały skonstruowane w celu użycia zarówno w warunkach klinicznych (ADL, IADL, Barthel Index, FSI), jak i w badaniach populacyjnych (ODCE DQ, MOS PFM), do wykorzystania w których z czasem zaadaptowano także skale ADL i IADL. W przypadku skal stosowanych w badaniach populacyjnych pomiar stanu funkcjonalnego odbywa się przez odwołanie się do subiektywnej oceny przez badanych stopnia ograniczeń w ich codziennej aktywności [23].

Stosunkowo nowym narzędziem do pomiaru stanu funkcjonalnego jest skala GARS (*Groningen Activity Restriction Scale*), która zawiera pytania wchodzące w skład stosowanych do tej pory — niezależnie od siebie i często równolegle — skal ADL oraz IADL [24]. Skala ADL opiera się na subiektywnej ocenie przez badanych ich niezależności w czynnościach o podstawowym znaczeniu w samoopiece, takich jak jedzenie, ubieranie się, utrzymywanie higieny ciała itp. [17, 24, 25]. Efektem takiej konstrukcji skali jest niedoszacowanie liczby osób potrzebujących pomocy, ponieważ nie zawiera ona informacji o indywidualnej zdolności dostosowania się do wymogów środowiska, w którym żyją badani. Dane te zawiera skala IADL, obejmująca pytania dotyczące możliwości samodzielnego poruszania się, robienia zakupów, prowadzenia domu itp. Zatem skala, która dostarczałaby wiarygodnych danych na temat zakresu upośledzenia funkcjonalnego badanych i wymaganej przez nich pomocy, powinna zawierać pytania należące zarówno do skali ADL, jak i IADL [18, 24]. Różne prace sugerują, że między zmiennymi należącymi do skal ADL oraz IADL, analizowanymi jako jeden zestaw pytań, istnieje zależność hierarchiczna, analogiczna do zależności obserwowanej w przypadku każdej ze skal oddzielnie [13, 24, 25]. Skala GARS jest próbą połączenia obu skal w jedno narzędzie, mające ich zalety, ale pozbawione ich wad [24].

Celem niniejszej pracy jest ocena trafności, rzetelności i homogeniczności skali GARS w populacji osób w starszym wieku w Polsce.

Material i metody

Ocenę parametrów psychometrycznych skali GARS przeprowadzono na podstawie wyników badania obejmującego losową próbę prostą 528 mieszkańców Krakowa w wieku 65–85 lat mieszkających we własnych gospodarstwach domowych. Średni wiek badanych wyniósł 72,7 roku; 59,5% badanych stanowiły kobiety, 21,9% respondentów miało wykształcenie pomaturalne i wyższe.

Polska wersja skali została utworzona na podstawie przekładu na język polski, dokonanego niezależnie przez dwóch tłumaczy, skali zawartej w kwestionariuszu wykorzystanym w holenderskich badaniach dotyczących jakości życia i stanu zdrowia osób starszych, w ramach projektu Nestor, prowadzonego przez *Northern Center for Health Care Research, University of Groningen*. Rozwinięciem tych badań były analizy prowadzone w ramach sieci *Central European Network*, obejmujące Polskę, Rumunię i Chorwację, w ramach których zebrano dane będące podstawą

niniejszego opracowania. Polska wersja kwestionariusza została następnie przetłumaczona z powrotem na język angielski, w celu określenia jej zgodności z wersją stosowaną w badaniach Nestor.

Skala GARS [24] zawiera pytania wchodzące w skład utworzonych wcześniej skal ADL [17] oraz IADL [18], powszechnie stosowanych do oceny stanu funkcjonalnego osób w starszym wieku, i jako połączenie tych dwóch skal zostanie przedstawiona analiza jej właściwości. Treść pytań i ich przynależność do poszczególnych podskal przedstawiono w tabeli 1.

Na każde pytanie wchodzące w skład skali respondent mógł udzielić następujących odpowiedzi: tak, mogę to robić w pełni samodzielnie bez żadnych trudności (ranga 1); tak, mogę to robić w pełni samodzielnie, ale z pewnymi trudnościami (ranga 2); tak, mogę to robić w pełni samodzielnie, ale z poważnymi trudnościami (ranga 3); nie mogę tego robić w pełni samodzielnie — mogę to robić jedynie z czyjąś pomocą (ranga 4). W celu zbadania trafności teoretycznej cała skala oraz poszczególne podskale wchodzące w jej skład zostały poddane analizie czynnikowej metodą głównych składowych. Oczekiwano, że zmienne tworzące skalę będą korelowały na poziomie co najmniej 0,7 (niektórzy badacze dopuszczają wartość 0,6) z pierwszą główną składową, reprezentującą teoretyczny konstrukt, który mierzy skala. Taki wynik wskazywałby na jednowymiarową strukturę skali. Następnie dokonano analizy rzetelności metodą badania wewnętrznej spójności skali za pomocą współczynnika alfa Cronbacha [26]. Wymaga się, aby badane skale charakteryzowały się wartością współczynnika alfa Cronbacha wyższą niż 0,7 (warunek ten nosi nazwę kryterium Nunnally'ego [27, 28]) oraz aby usuwanie zmiennych ze skali powodowało spadek wartości współczynnika. Oczekuje się także, że korelacje poszczególnych zmiennych z sumarycznym wynikiem skali będą wyższe niż 0,4 [29]. Do zbadania homogeniczności skali wykorzystano stochastyczny kumulatywny model skalowania Mokkena [30]. Na jego podstawie szacuje się następujące wskaźniki: współczynnik homogeniczności H Loewinger będący miarą homogeniczności skali jako całości [15, 16] oraz rodzinę współczynników zgodności (w literaturze anglojęzycznej określanych także jako *scalability coefficients*) H_i oceniających dopasowanie każdej ze zmiennych do skali, traktowanej jako model hierarchicznych zależności między zmiennymi [31, 32]. Wymaga się, aby współczynnik H oraz wszystkie współczynniki H_i były wyższe od pewnej stałej c ustalonej przez badacza, jednak sugeruje się, aby nie była ona niższa niż 0,3. Jeśli współczynniki osiągną wartość powyżej 0,5, uważa się, że zależ-

Tabela 1. Treść pytań tworzących skale wchodzące w skład skali GARS
Table 1. Wording of items constituting the GARS scale

Nr pytania	Treść pytania
ADL1	Respondent może chodzić po schodach
ADL2	Respondent może wychodzić z domu
ADL3	Respondent może się umyć i osuszyć całe ciało
ADL4	Respondent może obcinać paznokcie u nóg i pielęgnować stopy
ADL5	Respondent może się ubrać
ADL6	Respondent może się położyć i wstać z łóżka
ADL7	Respondent może wstać z krzesła i usiąść na krześle
ADL8	Respondent może umyć twarz i ręce
ADL9	Respondent może usiąść na muszli w toalecie i wstać z niej
ADL10	Respondent może jeść
ADL11	Respondent może poruszać się po domu (z laską, jeśli to konieczne)
IADL1	Respondent może przygotować śniadanie lub mały posiłek
IADL2	Respondent może przygotować obiad
IADL3	Respondent może wykonywać cięższe czynności domowe (np. mycie podłogi lub okien)
IADL4	Respondent może wyprać i wyprasować swoje ubrania
IADL5	Respondent może pościelić łóżko
IADL6	Respondent może robić zakupy
IADL7	Respondent może wykonywać lekkie czynności domowe (np. wycieranie kurzu, układanie rzeczy)

ność między zmiennymi tworzącymi skalę jest silnie hierarchiczna, dla współczynników należących do przedziału 0,4–0,5 mówi się o średniej, natomiast poniżej 0,4 — o słabej zależności hierarchicznej [15, 30, 33–35]. Obliczono także wartość współczynnika rzetelności rho, będącego odpowiednikiem współczynnika alfa Cronbacha stosowanego w przypadku skalowania Likerta i skal opartych na klasycznej teorii testu [9, 12, 36].

Wyniki

Analiza głównych składowych dla zmiennych wchodzących w skład podskali ADL skali GARS wykazała w warunkach polskich istnienie dwóch głównych składowych o wartościach własnych powyżej 1, z których pierwsza tłumaczy 59% wariancji zestawu zmiennych tworzącego skalę, podczas gdy druga tylko

13%. Korelacje wszystkich, poza dwiema, zmiennych tworzących podskalę z pierwszą z wyodrębnionych składowych są wyższe niż 0,7 (najniższa korelacja wynosi 0,58), co wskazuje na jednowymiarową strukturę tej podskali (tab. 2). Obie wspomniane zmienne — możliwość samodzielnego chodzenia po schodach oraz samodzielnej pielęgnacji stóp — korelują z drugą składową na zbliżonym, lecz nieco niższym poziomie niż z pierwszą. Korelacje z drugą główną składową o podobnych wartościach jak w przypadku dwóch zmiennych wspomnianych powyżej można zaobserwować także w przypadku trzech kolejnych zmiennych, dotyczących możliwości: samodzielnego wychodzenia z domu, jedzenia oraz mycia twarzy i rąk (tab. 2).

Tylko w przypadku dwóch zmiennych dotyczących samodzielnej pielęgnacji stóp i samodzielnego

Tabela 2. Macierz nierotowanych głównych składowych oraz analiza rzetelności podskali ADL skali GARS
Table 2. Matrix of nonrotated principal components, and reliability analysis of the ADL subscale of the GARS scale

	Składowa		Korelacja pytanie–skala	Alfa po usunięciu pytania	Średnia ranga	H _i
	1	2				
ADL1	0,63	0,56	0,65	0,874	1,82	0,76
ADL2	0,71	0,48	0,72	0,866	1,39	0,71
ADL3	0,80		0,75	0,862	1,16	0,73
ADL4	0,58	0,48	0,58	0,886	1,55	0,62
ADL5	0,86		0,73	0,869	1,07	0,81
ADL6	0,77		0,72	0,865	1,20	0,71
ADL7	0,85		0,73	0,869	1,09	0,73
ADL8	0,77	-0,51	0,59	0,879	1,02	0,90
ADL9	0,85		0,75	0,868	1,09	0,74
ADL10	0,71	-0,56	0,52	0,882	1,02	0,96
ADL11	0,86		0,76	0,868	1,08	0,80

W tabeli ukryto ładunki czynnikowe o wartościach < 0,3

jedzenia obserwuje się wzrost wartości współczynnika alfa Cronbacha po usunięciu tych zmiennych ze skali. Rzetelność podskali ADL skali GARS szacowana współczynnikiem alfa wynosi 0,88, a korelacje wszystkich zmiennych z sumarycznym wynikiem skali są wyższe niż 0,4 (tab. 2).

Wyniki analizy Mokkena pokazują, że skala ADL cechuje się silną hierarchią między zmiennymi — współczynnik homogeniczności H wynosi 0,73, natomiast współczynnik rzetelności rho — 0,93. W tabeli 2 przedstawiono hierarchię zmiennych tworzących skalę według średniej rangi odpowiedzi udzielonych na dane pytanie, począwszy od pytań „najłatwiejszych” (dotyczących czynności wykonywanych przez większość badanych samodzielnie), oraz współczynniki zgodności dla zmiennych tworzących skalę — wszystkie osiągnęły wartość powyżej 0,6 (tab. 2). Zestaw zmiennych tworzących podskalę IADL skali GARS w odniesieniu do analizy głównych składowych charakteryzuje się lepszymi parametrami niż podskala ADL. Pierwszy z czynników wyodrębnionych przez procedurę tłumaczy 59% wariacji tej grupy zmiennych, a drugi 17%. Korelacje zmiennych tworzących podskalę z pierwszą główną składową są wyższe niż 0,7 dla wszystkich zmiennych z wyjątkiem jednej, dotyczącej samodzielnego wykonywania cięższych czynności domowych, która ma na tej składowej ładunek o wartości 0,68 (tab. 3). Powyższe wyniki

wskazują na silnie jednowymiarową strukturę podskali IADL skali GARS.

Zmienne tworzące podskalę wykazują odpowiednio wysokie (wyższe niż 0,4) korelacje z sumarycznym wynikiem skali, a wartość alfa Cronbacha, wynosząca dla tej podskali 0,85, nie rośnie przy usuwaniu ze skali żadnego z pytań, które ją tworzą (tab. 3).

Wyniki analizy Mokkena pokazują, że skala IADL cechuje się silną hierarchiczną strukturą. Współczynnik skalowania H Loewinger dla skali wynosi 0,72, a wartości współczynników zgodności dla poszczególnych zmiennych są wyższe niż 0,6 (tab. 3). Współczynnik rzetelności rho wynosi 0,90.

Analizując cały zestaw zmiennych tworzących skalę GARS, trzeba zauważyć, że spośród 18 zmiennych tworzących skalę 12 koreluje z pierwszą główną składową na poziomie wyższym niż 0,7. Spośród pozostałych zmiennych trzy, dotyczące możliwości samodzielnego jedzenia, chodzenia po schodach oraz samodzielnego pielęgnacji stóp, korelują z pierwszą główną składową na poziomie 0,62–0,68, a trzy kolejne, dotyczące możliwości samodzielnego przygotowania obiadu, wykonywania cięższych czynności domowych oraz prania i prasowania, korelują z tą składową na poziomie 0,57–0,58. Większość tych zmiennych to zmienne, które w trakcie analizy podskal obniżały ich parametry. Spośród tych zmiennych dwie — przygotowanie obiadu i wykonywanie cięższych

Tabela 3. Macierz nierotowanych głównych składowych oraz analiza rzetelności podskali IADL skali GARS
Table 3. Matrix of nonrotated principal components, and reliability analysis of the IADL subscale of the GARS scale

	Składowa		Korelacja pytanie–skala	Alfa po usunięciu pytania	Średnia ranga	H _i
	1	2				
IADL1	0,81	-0,47	0,82	0,62	1,10	0,89
IADL2	0,74		0,81	0,62	1,53	0,61
IADL3	0,68	0,58	0,82	0,64	2,28	0,72
IADL4	0,75	0,45	0,81	0,72	1,96	0,73
IADL5	0,78	-0,44	0,83	0,59	1,09	0,84
IADL6	0,76	0,32	0,84	0,69	1,63	0,67
IADL7	0,86	-0,32	0,83	0,70	1,15	0,86

W tabeli ukryto ładunki czynnikowe o wartościach < 0,3

czynności domowych — korelują silniej z drugą, a nawet trzecią składową niż z pierwszą (tab. 4). Składowe te tłumaczą odpowiednio 13 i 6 % wariacji całego zestawu zmiennych wobec 54% wariacji tłumaczonej przez pierwszą składową. Wyniki te pozwalają uznać skalę GARS za skalę jednowymiarową.

Analiza relacji między pozycjami skali a jej wynikiem sumarycznym wskazuje, że wszystkie zmienne spełniają warunek Kleine'a, wymagający, aby korelacja z wynikiem skali była wyższa niż 0,4 — najniższe korelacje można zaobserwować głównie w przypadku zmiennych, które w trakcie analizy podskal obniżały ich parametry. Tylko w przypadku dwóch z nich — dotyczących możliwości samodzielnego jedzenia oraz mycia twarzy i rąk — po usunięciu pytania ze skali nie występuje wyraźny spadek wartości alfa Cronbacha równego dla całej skali 0,92 (tab. 5).

Wyniki analizy Mokkena pokazują, że skala GARS jest rzetelną skalą z silną hierarchią zmiennych — współczynnik homogeniczności H Loevinger wynosi 0.70, natomiast współczynnik rzetelności rho — 0,95. Współczynniki zgodności H_i dla poszczególnych zmiennych mieszczą się w zakresie 0,54–0,96, wskazując na wysoką homogeniczność skali GARS (tab. 5).

Dyskusja

Zaprezentowane wyniki wskazują na wysoką wartość skali GARS jako narzędzia służącego pomiarowi stanu funkcjonalnego osób w wieku starszym w populacji polskiej. Struktura zmiennych jest zbliżona do przedstawionej przez twórców skali: trzy wyodręb-

Tabela 4. Macierz nierotowanych głównych składowych dla pozycji skali GARS

Table 4. Matrix of nonrotated principal components for the items of the GARS scale

	Składowa		
	1	2	3
ADL1	0,68	0,44	
ADL2	0,77	0,32	
ADL3	0,80		
ADL4	0,62	0,33	
IADL1	0,79		0,38
IADL2	0,58		0,60
IADL3	0,58	0,59	
IADL4	0,57	0,52	0,32
IADL5	0,83		
IADL6	0,72	0,48	
ADL5	0,84	-0,34	
ADL6	0,75		-0,36
ADL7	0,79		
ADL8	0,70	-0,54	
ADL9	0,80		
ADL10	0,63	-0,57	
ADL11	0,82		
IADL7	0,83		

W tabeli ukryto ładunki czynnikowe o wartościach < 0,3

Tabela 5. Analiza wewnętrznej spójności oraz współczynniki skalowalności pozycji skali GARS
Table 5. Internal consistency analysis and scalability coefficients for the items of the GARS scale

	Korelacja pytanie–skala	Alfa po usunięciu pytania	Średnia ranga	H _i
ADL10	0,45	0,919	1,02	0,96
ADL8	0,52	0,918	1,02	0,90
ADL5	0,69	0,915	1,07	0,83
ADL11	0,69	0,915	1,08	0,78
IADL5	0,71	0,914	1,09	0,80
ADL9	0,68	0,915	1,09	0,71
ADL7	0,64	0,915	1,09	0,69
IADL1	0,69	0,913	1,10	0,79
IADL7	0,77	0,911	1,15	0,79
ADL3	0,72	0,912	1,16	0,72
ADL6	0,69	0,913	1,20	0,65
ADL2	0,78	0,909	1,39	0,71
IADL2	0,57	0,916	1,53	0,54
ADL4	0,62	0,914	1,55	0,58
IADL6	0,76	0,910	1,63	0,72
ADL1	0,71	0,911	1,82	0,71
IADL4	0,64	0,916	1,96	0,66
IADL3	0,65	0,916	2,28	0,75

nione główne składowe tłumaczą odpowiednio: 54, 13 i 6 — łącznie 73% — całkowitej wariancji zestawu zmiennych, w porównaniu z, odpowiednio, 48, 11 i 6 — łącznie 65% — otrzymanych przez autorów skali [24]. Wysoki odsetek łącznej wariancji pozycji skali tłumaczony przez pierwszą główną składową oraz wysoki stosunek wariancji wyjaśnianej przez pierwszą składową do wariancji wyjaśnianej przez drugą składową pozwalają uznać strukturę czynnikową skali za jednowymiarową. Macierz czynników otrzymana w badaniu własnym charakteryzuje się większą przejrzystością niż zaprezentowana przez autorów skali — mniej jest zmiennych korelujących na zbliżonym poziomie z kilkoma czynnikami jednocześnie, a korelacje z pierwszą składową są wyższe [24]. Według wyników eksploracyjnej analizy czynnikowej przedstawionych przez Kempena [24], skala GARS ma strukturę trójwymiarową, z ukrytym jednym wymiarem wyższego poziomu, którego istnienie odzwierciedla

się w tym, że wiele zmiennych tworzących skalę ma porównywalne ładunki czynnikowe na więcej niż jednej składowej. Wyniki analizy czynnikowej bez rotacji, otrzymane przez autorów skali, pokazują, że na pierwszym wymiarze zmienne wchodzące w skład skali mają ładunki 0,42–0,79, przy czym 15 z nich ma ładunki wyższe niż 0,6 [24]. W badaniach własnych otrzymano ładunki czynnikowe na pierwszej głównej składowej w zakresie 0,57–0,84, przy czym również tylko trzy z nich były niższe od 0,6, co świadczy o silniejszym związku między poszczególnymi zmiennymi tworzącymi polską wersję skali oraz teoretycznym konstruktorem mierzonym przez narzędzie. Rzetelność skali mierzona współczynnikiem alfa Cronbacha jest równie wysoka jak otrzymana przez autorów skali: w badaniach własnych otrzymano wartość alfa Cronbacha na poziomie 0,92 (wobec 0,90 w oryginalnej pracy), podczas gdy w innych badaniach uzyskano współczynniki o wartościach: 0,94 [14, 37] oraz 0,91 [24].

Rzetelność skali mierzona współczynnikiem rho, analizowanym w przypadku modelu skalowania Mokkena, jest wyższa niż podana przez autorów skali: w badaniach własnych otrzymano wartość współczynnika równą 0,95, podczas gdy twórcy skali podają wartość 0,93 [24], inni zaś autorzy otrzymali wartości 0,90–0,96 [13, 14, 37].

Współczynniki skalowalności dla polskiej wersji wynosiły odpowiednio: dla całej skali $H = 0,70$ (autorzy skali otrzymali wartość 0,63) oraz dla poszczególnych zmiennych współczynniki H_i zawarte w przedziale 0,61–0,96 (wg autorów 0,51–0,75), podczas gdy inni autorzy otrzymali współczynnik skalowalności H dla całej skali o wartościach: 0,65 [14, 37] oraz 0,47 [13], a dla większości zmiennych w zakresach 0,56–0,76 (przy czym dwie zmienne charakteryzowały wartości 0,41 i 0,89) [14, 37] oraz 0,42–0,53 [13]. Używane do tej pory powszechnie skale ADL i IADL (jako podskale wydzielone ze skali GARS) w odniesieniu do analizy Mokkena w badaniach własnych charakteryzowały się podobnymi (lub wyższymi) wartościami współczynników oceniających ich walory psychometryczne w porównaniu z wartościami podanymi przez innych autorów. Rzetelność tych skal mierzona współczynnikiem alfa Cronbacha wyniosła w badaniach własnych dla skali ADL 0,88, a dla skali IADL 0,85, podczas gdy inni autorzy podają wartości, odpowiednio, 0,92 i 0,98 [14, 37]. Rzetelność podskal mierzona za pomocą współczynnika rho wyniosła w badaniach własnych dla skali ADL 0,93, natomiast dla skali IADL 0,90, podczas gdy inni autorzy podają wartości, odpowiednio, 0,95 i 0,93 [14, 37] oraz, odpowiednio, 0,90 i 0,89 [13].

Współczynniki homogeniczności H dla poszczególnych podskal wyniosły w badaniach własnych dla skali ADL 0,73, a dla skali IADL 0,72, podczas gdy inni autorzy podają wartości, odpowiednio, 0,69 i 0,70 [14, 37]

oraz, odpowiednio, 0,52 i 0,51 [13]. Współczynniki homogeniczności H_i dla zmiennych tworzących poszczególne podskale w badaniach własnych mieściły się w zakresie, odpowiednio, dla skali ADL 0,62–0,96 oraz dla skali IADL 0,61–0,89, Kempen zaś podawał wartości 0,51–0,77 dla skali ADL oraz 0,63–0,91 dla skali IADL [14].

Wyniki podawane przez autorów prac wydanych przed 1996 rokiem nie dają się jednak w pełni porównać z wynikami badań własnych ani z wynikami Kempena [24], gdyż dotyczą wcześniejszych wersji skali GARS z odpowiedziami należącymi do trzech kategorii [14, 37], w przypadku których analizy były dokonywane za pomocą stochastycznego kumulatywnego modelowania Mokkena dla zmiennych dychotomicznych, po uprzednim przekodowaniu zmiennych do dwóch wartości, ewentualnie dla wersji o pięciu kategoriach odpowiedzi, w przypadku której analizy były dokonywane także za pomocą modelu Mokkena dla zmiennych politomicznych [13, 14, 37].

Otrzymane wyniki pozwalają uznać skalę GARS za trafne, rzetelne oraz homogeniczne narzędzie, o silnej hierarchii zmiennych, służące ocenie stanu funkcjonalnego osób w starszym wieku w Polsce.

Wnioski

- Skala *Groningen Activity Restriction Scale* zastosowana w populacji osób w starszym wieku w Polsce cechuje się wysoką trafnością.
- Skala GARS jest skalą jednowymiarową cechującą się silną homogenicznością.
- Rzetelność skali w polskiej populacji osób starszych jest wyższa niż oszacowana przez jej twórców.

Podziękowania

Praca finansowana ze środków grantu 501/PKL/99/L.

Streszczenie

Wstęp. Zastosowanie skali badawczej opracowanej poza Polską w próbie osób starszych mieszkających w Polsce wymaga zbadania jej trafności i rzetelności. Jest to konieczne ze względu na możliwość istnienia różnic w znaczeniu lub rozumieniu pytań tworzących skalę mierzącą jakość życia warunkowaną stanem zdrowia. Taka procedura jest niezbędna również w przypadku skali mierzącej stan funkcjonalny, choć wydaje się, że sformułowania w poszczególnych pozycjach takiej skali nie odwołują się do wyrażen i symboli silnie odnoszących się do elementów charakterystycznych dla kultury kraju, w którym skala została opracowana.

Materiał i metody. Przedstawione wyniki pochodzą z badania przeprowadzonego w losowej próbie prostej 528 mieszkańców Krakowa w wieku 65–85 lat zamieszkałych we własnych gospodarstwach domowych. Analizę właściwości psychometrycznych skali GARS w kontekście klasycznej teorii testu przeprowadzono za pomocą analizy czynnikowej (trafność) oraz współczynnika alfa Cronbacha (rzetelność w aspekcie wewnętrznej spójności). Analizy właściwości psychometrycznych skali w kontekście teorii odpowiedzi na pozycje testu dokonano za po-

mocą procedury skalowania Mokkena: oszacowano wartość współczynnika rzetelności ρ oraz współczynnika homogeniczności H Loevinger.

Wyniki. Wszystkie pozycje skali GARS korelowały z pierwszą główną składową na poziomie wyższym niż 0,56, co wskazuje na jednowymiarową strukturę skali. Skala GARS charakteryzowała się wysoką rzetelnością, zarówno w kontekście klasycznej teorii testu ($\alpha = 0,92$), jak i w kontekście teorii odpowiedzi na pozycje testu ($\rho = 0,95$). Współczynnik homogeniczności H dla skali GARS wynosił 0,70, dla poszczególnych pozycji zaś zawierał się w granicach 0,54–0,96.

Wnioski. Skala GARS charakteryzuje się wysoką trafnością i rzetelnością oraz silną homogenicznością.

Gerontol. Pol. 2012; 20, 3: 109–117

Słowa kluczowe: *trafność, rzetelność, homogeniczność, zgodność, jakość życia warunkowana stanem zdrowia, GARS, stan funkcjonalny, osoby w starszym wieku*

Piśmiennictwo

1. Verbrugge L. Disability in late life. W: Abeles R.P., Gift H.C., Ory M.G. (red.). Aging and quality of life. Springer Publishing Company, Nowy Jork 1997; 179–198.
2. Donaldson L.J., Clayton D.G., Clarke M. The elderly in residential care: mortality in relation to functional capacity. *J. Epid. Com. Health* 1980; 2: 96–101.
3. Koyano W., Shibata H., Nakazato K. Mortality in relation to instrumental activities of daily living: one-year follow-up in a Japanese urban community. *Soc. Sci.* 1989; 3: 107–109.
4. Grand A., Grosclaude P., Bocquet H., Pous J., Albaredo J.L. Disability, psychosocial factors and mortality among the elderly in a rural French population. *J. Clin. Epidemiol.* 1990; 8: 773–782.
5. Tobiasz-Adamczyk B. Czynniki psychospołeczne warunkujące długość życia osób starszych wiekiem. *Gerontol. Pol.* 1997; 1: 30–37.
6. Tobiasz-Adamczyk B., Zawisza D. Relacja między wsparciem otrzymywanym w ramach interakcji społecznych a ryzykiem zgonu w populacji osób starszych w Krakowie. W: Kowaleski J.T., Szukalski P. (red.). Pomyślne starzenie się w świetle nauk o zdrowiu. Zakład Demografii i Gerontologii Społecznej Uniwersytetu Łódzkiego, Łódź 2008; 143–153.
7. Tobiasz-Adamczyk B., Brzyski P. Czynniki warunkujące zmiany w stanie funkcjonalnym w wieku starszym na podstawie 12-letniej obserwacji. *Gerontol. Pol.* 2002; 10: 23–29.
8. Hornowska E., Paluchowski W.J. Kulturowa adaptacja testów psychologicznych. W: Brzeziński J. (red.). Metodologia badań psychologicznych. Wybór tekstów. Wydawnictwo Naukowe PWN, Warszawa 2004; 151–191.
9. Brzeziński Z. Metodologia badań psychologicznych. Wydawnictwo Naukowe PWN, Warszawa 2006.
10. Brzyski P. Trafność i rzetelność skal pomiarowych. W: Kawecka-Jaszcz K., Kłoczek M., Tobiasz-Adamczyk B. (red.). Jakość życia w chorobach układu sercowo-naczyniowego. Metody pomiaru i znaczenie kliniczne. Termedia, Poznań 2006; 43–54.
11. McDowell I. Measuring health: a guide to rating scales and questionnaires. Wyd. 3. Oxford University Press, Oxford 2006.
12. Gulliksen H. Theory of mental tests. Wiley, Nowy Jork 1950.
13. Suurmeijer T.P.B.M., Doeglas D.M., Moum T. i wsp. The Groningen Activity Restriction Scale for measuring disability: its utility in international comparison. *Am. J. Publ. Health* 1994; 84: 1270–1273.
14. Kempen G.I.J.M., Suurmeijer T.P.B.M. The development of a hierarchical polytomous ADL-IADL scale for noninstitutionalized elders. *The Gerontological Society of America* 1990.
15. Loewinger J. A systematic approach to the construction and evaluation of tests of ability. *Psychological Monographs* 1947; 4: 1–49.
16. Guilford J.P. Teoria testów psychologicznych. W: Brzeziński J. (red.). Trafność i rzetelność testów psychologicznych. Wybór tekstów. Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2005; 26–61.
17. Katz S., Ford A.B., Moskowitz R., Jackson B.A., Jaffe M.W. Studies of illness in the aged: the index of ADL: a standardised measure of biological and psychosocial function. *JAMA* 1963; 185: 914–919.
18. Lawton M.P., Brody E. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist* 1969; 9: 179–186.
19. Mahoney F.I., Barthel D.W. Functional evaluation: the Barthel Index. *Md State Med. J.* 1965; 14: 61–65.
20. McWinnie J.R. Disability assessment in population surveys: results of the OECD common development effort. *Rev. Epidemiol. Sante Publique* 1981; 29: 413–419.
21. Jette A.M. Functional capacity evaluation: an empirical approach. *Arch. Phys. Med. Rehabil.* 1980; 61: 85–89.
22. Stewart A.L. Physical functioning measures. W: Stewart A.L., Ware J.E. Jr (red.). Measuring functioning and well-being: The Medical Outcomes Study approach. Duke University Press, Durham 1992; 86–101.
23. Bowling A. Measuring health. A review of quality of life measurement scales. Open University Press, Buckingham, Filadelfia 1997.
24. Kempen G.I.J.M., Miedema I., Ormel J., Molenaar W. The assessment of disability with the Groningen Activity Restriction Scale. Conceptual framework and psychometric properties. *Soc. Sci. Med.* 1996; 11: 1601–1610.
25. Spector W.D., Katz S., Murphy J.B., Fulton J.P. The hierarchical relationship between activities of daily living and instrumental activities of daily living. *J. Chron. Conditions* 1987; 6: 481–489.
26. Cronbach L.J. Współczynnik alfa a struktura wewnętrzna testów. W: Brzeziński J. (red.). Trafność i rzetelność testów psychologicznych. Wybór tekstów. Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2005; 177–212.
27. Nunnally J.C., Bernstein I.H. Psychometric theory. Wyd. 3. McGraw-Hill, Nowy Jork 1994.
28. Bowling A. Research methods in health: investigating health and health services. Open University Press, Buckingham, Filadelfia 2002.
29. Kline P. A handbook of test construction. Methuen, Londyn 1986.
30. Mokken R.J. A theory and procedure of scale analysis. The Hague, Moulton 1971.
31. Loewinger J. The technic of homogeneous tests compared with some aspects of scale analysis and factor analysis. *Psychol. Bull.* 1948; 19: 507–529.
32. Guilford J.P. Tworzenie testu. W: Brzeziński J. (red.). Trafność i rzetelność testów psychologicznych. Wybór tekstów. Gdańskie Wydawnictwo Psychologiczne, Gdańsk 2005; 107–166.
33. Mokken R.J., Lewis C. A non-parametric approach to the analysis of dichotomous item responses. *Appl. Psychol. Measur.* 1982; 6: 417–430.
34. De Jong A., Molenaar W. An application of Mokken's model for stochastic cumulative scaling in psychiatric research. *J. Psychiatr. Res.* 1987; 21: 137–149.
35. Molenaar W., Debets P., Sijtsma K., Hemker B.T. MSP, a program for Mokken Scale analysis for Polytomous items, user's manual, version 3.0. iecProgramma, Groningen 1994.
36. Molenaar W. Mokken Scaling revisited. *Kwantitatieve methoden* 1982; 3: 145–164.
37. Suurmeijer T.P.B.M., Kempen G.I.J.M. Behavioural changes as an outcome of disease: the development of an instrument. *Int. J. Health Sciences* 1990; 1: 189–194.