

Małgorzata Rzewuska

Badanie zgodności i niezawodności polskiej wersji skali PANSS

Samodzielna Pracownia Farmakoterapii Instytutu Psychiatrii i Neurologii w Warszawie

Streszczenie

W pracy przedstawiono wyniki treningu oceny stanu psychicznego w skali PANSS. Oceniono zgodność ocen 418 lekarzy oraz porównano uzyskane oceny ze standardową. Porównanie pozwoliło na sformułowanie do niektórych objawów skali komentarzy, które mogą korzystnie wpłynąć na większą zgodność ocen ze standardową.

Wprowadzenie

Skala oceny zaburzeń psychicznych w schizofrenii – PANSS powstała w 1987 r. Autorzy skali (Key, Oper i Lindenmayer) utworzyli ją w oparciu o szeroko stosowaną od 1962 r. w badaniach klinicznych 18 punktową BPRS (Overall i Gorham 1988) oraz PRS (Psychopathology Ratings Scale) zawierającą objawy negatywne i świadczące o zaburzeniach funkcjonowania społecznego. Skala PANSS pozwala na ocenę 30 cech zawartych w trzech podskalach: objawów pozytywnych, negatywnych i ogólnych. Skala ta jest obecnie podstawowym narzędziem do oceny dynamiki zaburzeń w badaniach klinicznych działania leków.

Zarówno jej wersja oryginalna, jak i kilka wersji językowych (Knorrig i wsp. 1992, Müller i wsp. 1998, Lindström i wsp. 1994) zostało ocenione co do niezawodności i pewności (zgodności ocen). W Polsce od wielu lat używa się tłumaczeń skali, głównie w międzynarodowych wielośrodkowych badaniach nowych leków przeciwpsychotycznych. Wielu lekarzy uczestniczących w badaniach przeżyło trening w ocenie skali PANSS prowadzony przy pomocy oryginalnych nagrań video badania z interpretacją autorów skali. Uczestnicząc wielokrotnie w treningach międzynarodowych z udziałem polskich psychiatrów dostrzegłam, że nasze oceny są stale nieco odmienne od oceny normatywnej w kilku punktach skali. Dotyczy to głównie punktów, które dotyczą wypowiedzi świadczących o przeżywaniu subiektywnym (depresja, niepokój) oraz niektórych punktów których nasilenie uzależniono od urojeniowego przeżywania siebie (np. troska o zdrowie). W tradycji polskiej psychiatrii, ściśle związanej z psychopatologią niemiecką (Jaspers, K. Schneider), w której

na ocenę objawów wpływają inne, a nie sama tylko obserwacja (i dosłowne wypowiedzi) cechy, lekarze psychiatrzy zdają się kierować bardziej osądem związku wypowiedzi z ich ekspresją niż tylko z treścią.

Obserwacje te skłoniły mnie do przeprowadzenia ocen skali PANSS z udziałem dużej liczby psychiatrów.

Celem pracy było zbadanie zgodności ocen stanu psychicznego w polskiej wersji skali PANSS i porównanie ich z ocenami oryginalnymi. Ustalenie przyczyn rozbieżności mogłoby pozwolić na opracowanie komentarza który podkreśliłby intencje autorów skali.

Metoda

Do oceny użyto oryginalnego nagrania badania jednego pacjenta. Przed wyświetleniem badania ekspert omówił poszczególne punkty skali zgodnie z definicjami podanymi w podręczniku. Oceny dokonywało 418 psychiatrów (w 22 grupach liczących po 17–25 uczestników). Każde badanie zakończono analizą wzajemnej zgodności ocen uczestników oraz omówiono różnice w porównaniu z oceną standardową. Strukturę posiedzeń przedstawiono w tabeli 1.

Tabela 1. Struktura posiedzenia

wprowadzenie	10'
Video – PANSS	40–45'
wypełnianie skali	30–40'
porównanie ocen, dyskusja	60–90'

Do oceny zgodności badaczy stosowano metodę zaproponowaną przez Kay'a (1991).

Za ocenę zgodną uznano tę, w której punktacja nie odbiegała od oceny standardowej (oryginalnej) o więcej niż o 1 punkt. Liczbę prawidłowych ocen podsumowano w podskalach (pozytywnej, negatywnej i ogólnej) oraz całościowo. Według danych z piśmiennictwa zgodność w 70% można uznać „do zaakceptowania” zaś w 90% jest „duża”. Jeśli 80% badających postawi 80% zgodnych ocen, wynik treningu jest uważany za zadowalający (Kay i in. 1988, Lambert 1996, Lindström i in. 1994, Knorring i Lindström 1992). Do oceny zgodności ze standardem użyto testu Cochrańa – Q, jako poziom istotności uznano poziom $\alpha=0,05$.

Dodatkowo obliczono wskaźnik zgodności Cohena (Kw kappa) do oceny odpowiedzi poprawnych z błędem ± 1 punkt.

Uwzględniono także obliczenie przypadkowej zgodności ocen przy pomocy współczynnika ICC mierzącego statystyczną istotność zgodności ocen. Za zgodne przyjęto oceny gdy wartość współczynnika ICC wyniósł ponad $>0,90$ natomiast za zadowalające, w których ICC wyniósł $0,40-0,50$.

Omówienie wyników

1. Analiza zgodności ocen wewnątrz grupy badających

Średnie oceny w podskalach objawów pozytywnych, negatywnych i ogólnych były do zaakceptowania, zgodność oceny (± 1) wyniosła dla podskali objawów pozytywnych 76%, negatywnych 81,2%, ogólnych 73,8% zaś w całej skali 76%.

Zgodność pomiędzy badającymi w przypadku 23 objawów (77%) była dobra. Ponad 90% lekarzy tak samo oceniło 9 objawów, a ponad 80% badających podobnie punktowało 14 objawów. Zgodność ocen zestawiono w tabeli 2.

Mniejsza nieco grupa (77–79%) zgodnie oceniła dalszych 5 objawów.

Najmniejsza zgodność dotyczyła oceny nasilenia zmanierowania (71%) oraz troski o zdrowie (65%).

Tabela 2. Zgodność ocen w grupie badających

A. zgodność ocen bardzo dobra

Zgodność ocen w grupie badających (N=418)

Objaw	Odsetek zgodnych ocen	Kappa ocena rozrzutu punktacji
N2 Emocjonalne wycofanie się	94	0,94
P2 Formalne zaburzenia myślenia	93	0,93
G9 Niezwykłe treści myślowe	93	0,29
N6 Mowa aspontaniczna/urywana	92	0,91
P1 Urojenia	91	0,17
G4 Napięcie	91	0,21
G13 Zaburzenia woli	91	0,13
N4 Bierność/apatia społeczne wycofanie	90	0,35
N5 Zaburzenia myślenia abstrakcyjnego	90	0,99

B. zgodność ocen dobra

Objaw	% zgodnych ocen	Kappa
P6 Podejrzliwość/prześladowanie	89	0,98
N7 Myślenie stereotypowe	89	0,14
G2 Lęk	89	0,14
G10 Dezorientacja	89	0,97
P3 Omamy	87	0,98
G11 Zaburzenia uwagi	87	0,99
G12 Brak krytycyzmu	85	0,26
N3 Utrudniony kontakt/Porozumienie	84	0,98
P7 Wrogość	83	0,44
G14 Drażliwość	82	0,95
P5 Postawa wielkościowa	81	0,21
G15 Zaabsorbowanie własnymi myślami	81	0,95
G16 Czynne unikanie kontaktów społecz.	80	0,96
G3 Poczucie winy	80	0,10

c.d. tab. 2

C. Oceny objawów najmniej zgodne

N1 Błady afekt	79	0,94
G6 Depresja	79	0,30
P4 Pobudzenie	78	0,43
G8 Brak współpracy	78	0,11
G7 Zahamowanie ruchowe	77	0,37
G5 Manieryzmy i podstawy	71	0,18
G1 Troska o zdrowie fizyczne	65	0,96

Wynik ten może świadczyć albo o trudności i ocenie tych objawów, albo, co wydaje się bardziej prawdopodobne z powodu dużego rozrzutu ocen, z różną interpretacją punktów skali.

2. Analiza zgodności ocen z podanym standardem

W porównaniu do zgodności ocen w obrębie grupy oceniających, zgodność ocen ze standardową była mniejsza.

Zgodność ocen 90% respondentów dotyczyła 20% objawów, a blisko połowę objawów oszacowało prawidłowo (w stosunku do standardu) mniej niż 80% oceniających.

Ranking zgodności ze standardem dla poszczególnych objawów przedstawiono w tabeli 3 (oceny zbliżone) i tabeli 4 (oceny rozbieżne).

Tabela 3. Oceny zbliżone ze standardową

Objawy	zgodna w grupie (A)	zgodna ze standardową (B)	różnica A-B	Kappa (A v/s B)	ICC (rozrzut)
P2 formalne zab. myślenia	93	93	0	0,93	0,93
P3 omamy	87	87	0	0,98	0,98
P6 podejrzliwość	89	89	0	0,98	0,97
N1 błady afekt	79	79	0	0,94	0,94
N2 wycofanie emocjonalne	94	94	0	0,94	0,94
N3 zaburzenia kontaktu	84	84	0	0,98	0,98
N5 myślenie abstrakcyjne	90	90	0	0,99	0,98
N6 aspontaniczność mowy	92	92	0	0,91	0,91
G9 niezwykle treści myślenia	93	90	3	0,29	0,30
G10 dezorientacja	89	84	5	0,97	0,97
G11 zaburzenia uwagi	87	87	0	0,99	0,99
G14 drażliwość	82	91	1	0,95	0,95
G15 zaabsorbowanie myślami	81	81	0	0,95	0,95
G4 napięcie	91	91	0	0,21	0,28
G16 kontakty społeczne	80	80	0	0,98	0,98

Tabela 4. Oceny odległe od oceny standardowej

Objaw	% zgodności w grupach badających A	% zgodności z oceną normatywną B	A-B	Kappa	ICC
P7 wrogość	83	72	11	0,44	0,46
P4 pobudzenie	78	65	13	0,43	0,44
G6 depresja	79	62	17	0,30	0,40
N4 bierność	90	81	9	0,35	0,36
G12 brak krytycyzmu	85	76	9	0,26	0,27
P5 postawa wielkościowa	81	71	10	0,21	0,22
G1 troska o zdrowie	65	55	10	0,96	0,93
G5 manieryzmy	71	81	10	0,18	0,24
P1 urojenia	91	56	35	0,17	0,24
N7 myślenie stereotypowe	89	49	40	0,14	0,20
G2 lęk	89	77	12	0,14	0,20
G13 zaburzenia woli	91	57	34	0,13	0,17
G8 brak współpracy	78	71	7	0,11	0,12
G3 poczucie winy	80	51	29	0,10	0,16

Tabela 5. Większa zgodność wśród badających niż z oceną standardową

Objaw	% zgodnych ocen badających A	% ocen zgodnych ze standardową B	różnica A-B	ICC
P1 urojenia	91	56	35	0,24
G13 zaburzenie woli	91	57	34	0,17
N7 stereotypia myślenia	89	49	40	0,20
G3 poczucie winy	80	51	29	0,16
G6 depresja	79	62	17	0,40
P7 wrogość	83	72	11	0,46
P5 postawa wielkościowa	81	71	10	0,22
N4 bierność	90	81	9	0,36
G12 brak krytycyzmu	85	76	9	0,27

Większa zgodność z oceną standardową niż pomiędzy badającymi

G5 manieryzmy	71	81	10	0,24
---------------	----	----	----	------

3. Analiza punktów skali, które oceniano za wysoko i za nisko w porównaniu do wzorca (por. Anaks)

Punkty w skali oceniane za wysoko i za nisko przedstawiono w tabeli 6.

Dyskusja

Badanie wykazało, że istnieje szereg różnic w ocenie stanu psychicznego dokonywanych przez psychiatrów polskich i przez autorów skali PANSS. Różnice ocen w obrębie grupy badających są znacznie mniejsze, a także spełniają kryteria pozwalające na uznanie ich za dostatecznie zgodne.

Tabela 6. Największe różnice ocen w porównaniu do wzorca (% odpowiedzi)

Oceny wyższe od standardowej

	+1	+2	> +2	łącznie
P1 urojenia	24	36	8	68
P4 pobudzenie	47	18	17	82
N7 myślenie stereotypowe	30	42	9	81
G 13 zaburzenie woli	44	36	6	86

Oceny niższe od standardowych

	-1	-2	-2	łącznie
G1 troska o zdrowie	11	16	29	56
G3 poczucie winy	23	36	11	70
G6 manieryzmy	34	27	11	72
G7 depresja	37	27	15	79
N4 bierność	33	12	6	51

Zarówno porównanie wyników, jak i oceny zarejestrowane w czasie dyskusji po treningach pozwalają na następujące komentarze, wyjaśniające rozbieżności w punktacji skali:

Urojenia (P1)

Wyższe oceny nasilenia urojeń postawiło 68% oceniających. Wynikały one przede wszystkim z przesłanki, że związki i myśli urojeniowe w znacznym stopniu określały zachowania pacjenta i relacje z otoczeniem. Tradycyjnie w polskiej psychiatrii systematyzacja urojeń nie jest postrzegana jako czynnik oznaczający większy stopień zaburzeń. Toteż wydaje się, że punkt powinien w polskiej wersji skali być opatrzony komentarzem, że przy ocenie niezbędne jest uwzględnienie systematyzacji treści, przyjętej za miarę nasilenia.

Zaburzenia w sferze woli (G13)

Zaburzenia w sferze woli były oceniane przez polskich psychiatrów wyżej prawdopodobnie „kosztem” niższej oceny bierności i apatii. Wydaje się, że obydwie punkty powinny być nieco dokładniej sprecyzowane, gdyż bierność i brak zainteresowania może być oceniane jako niezdecydowanie. Na różnicę ocen w tych punktach wpływa trudność w zrozumieniu wypowiedzi (treść była tłumaczona w napisach, ale nie wyraziła ona różnicy).

Depresja (G6), Poczucie winy (G3)

Duża różnica dotycząca znacznie niższego oceniania przez Polaków depresji (G6) i poczucia winy (G3), wynikała, co podkreślano wielokrotnie w dyskusji, z tego że autorzy skali większe znaczenie przydawali deklaracjom słownym, natomiast badający uwzględniali brak depresyjnej ekspresji w czasie badania. Te punkty skali warto opatrzyć komentarzem, że uwzględniać należy przede wszystkim treść wypowiedzi, a nie wrażenie badającego.

Troska o zdrowie fizyczne (G1)

Różnica w ocenie punktu G1-troska o zdrowie fizyczne wynikała stąd, że urojenia pacjenta miały charakter absurdałny. W tym punkcie niezbędny jest komentarz wskazujący na to, że ocenia się tu urojenia zmiany ciała niezależnie od ich prawdopodobieństwa, jako hipochondryczne

W polskim badaniu, podobnie jak w innych opisanych treningach (Andreasen i in. 1982, Müller i in. 2000, Peratla i Cuesto 1994) kwalifikacje zawodowe (długi staż w zawodzie, stopień specjalizacji) nie wpływały na poprawność ocen.

Wnioski

Przeprowadzone badanie wykazało różnice ocen kilku punktów skali wymagające dodatkowego komentarza do polskiej wersji skali PANSS. Dotyczą one różnic w ocenach nasilenia patologii. Uwagi powinny dotyczyć następujących objawów:

P1 urojenia

G1 troska o zdrowie fizyczne

G3 poczucie winy

G6 depresja

Różnice w ocenach pozostałych objawów mogą ulec zmniejszeniu po wtórzeniu przez badających treningów i moim zdaniem nie wymagają dodatkowego komentarza do podręcznika skali.

Dziękuję firmie Janssen-Cilag, a zwłaszcza dr med. Alinie Kućmie za przygotowanie i udostępnienie nagrań i organizację treningów.

Piśmiennictwo

1. Andreasen N.C.: Negative symptoms in schizophrenia: definition and reliability. *Arch. Gen. Psychiatry*, 1982, 39, 784–788
2. Kay S.R.: Positive and negative syndromes in schizophrenia. Brunner & Mazel, New York, 1991
3. Kay S.R.: Positive-negative symptom assessment in schizophrenia: psychometric issues and scale comparison. *Schizophr. Bull.*, 1987, 13, 261–276
4. Lambert T.: Inter-rater reliability and the PANSS. Poster presented at the CINP Congress, Melbourne, Australia, 1996
5. Landis J.R., Koch G.G.: The measurement of observer agreement for categorical data. *Biometrics*. 1977, 33, 159–174
6. Lindström E., Wieselgren I.M., von Knorring L.: Interrater reliability of the structured clinical interview for the Positive and Negative Syndrome Scale for schizophrenia. *Acta Psychiatr. Scand.*, 1994, 89, 192–195
7. Müller M.J., Rossbach W., Dannigkeit P., Müller-Siecheneder F., Szegedi A., Wetzell H.: Evaluation of standardized rater training for the Positive and Negative Syndrome Scale (PANSS). *Schizophr. Res.*, 1998, 32, 151–160

8. Müller M.J., Rossbach W., Davids E., Wetzel H., Benkert O.: Evaluation eines standardisierten trainings für die Positive and Negative Syndrome Scale (PANSS). *Nervenarzt.* 2000, 71, 195–204
9. Overall J.E., Gorham D.R.: The Brief Psych/hiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacol. Bull.*, 1988, 24, 97–99
10. Peralta V., Cuesta M.J.: Psychometric properties of the Positive and Negative Syndrome Scale (PANSS) in Schizophrenia. *Psychiatry Res.*, 1994, 53, 31–40
11. Singh M.M., Kay S.R.: A comparative study of haloperidol and chlorpromazine in terms of clinical effects and therapeutic reversal with benzotropine in schizophrenia: theoretical implications for potency differences among neuroleptics. *Psychopharmacologia*, 1975, 43, 103–113
12. Von Knorring L., Lindström E.: The swedish version of the Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Construct validity and interrater reliability. *Acta Psychiatr. Scand.*, 1992, 86, 463–468

Aneks

Odsetek ocen nasilenia poszczególnych objawów w skali PANSS

1. Objawy pozytywne

Nasilenie objawów w punktach	1	2	3	4	5	6	7
Urojenia	0	0	1	31	24	36	8
Formalne zaburzenia myślenia	0	0	4	19	46	29	2
Omamy	0	1	4	7	42	36	10
Pobudzenie	5	13	48	19	15	2	0
Postawa wielkościowa	7	20	36	25	10	2	0
Podjęzliwość/prześladowanie	0	1	6	9	49	31	4
Wrogość	5	19	49	16	8	3	0

Ocenę normatywną zacięniowano

2. Objawy negatywne

Nasilenie objawów w punktach	1	2	3	4	5	6	7
Błady afekt	0	1	17	19	40	20	3
Emocjonalne wycofanie się	0	1	3	11	46	36	3
Utrudniony kontakt/porozumienie	0	1	14	50	20	15	0
Bierność/apatia, społeczne wycof.	0	1	5	12	34	44	4
Zaburzenia myślenia abstrakcyjn.	0	1	4	19	27	44	5
Mowa aspontaniczna/urywana	0	0	6	48	39	7	0
Myślenie stereotypowe	0	2	17	30	42	8	1

Ocenę normatywną zacięniowano

3. Objawy ogólne

Nasilenie objawów w punktach	1	2	3	4	5	6	7
Troska o zdrowie fizyczne	6	23	16	11	38	6	0
Lęk	3	16	36	37	4	4	0
Poczucie winy	4	7	36	23	21	7	2
Napięcie	2	5	39	32	20	2	0
Manieryzmy i postawy	1	10	27	33	20	8	1
Depresja	4	11	26	37	16	6	0
Zahamowanie ruchowe	11	11	26	40	11	1	0
Brak współpracy	4	14	17	40	21	4	0

Ocenę normatywną zacięniowano

4. Objawy ogólne cd.

Nasilenie objawów w punktach	1	2	3	4	5	6	7
Niezwykłe treści myślowe	0	0	2	4	44	43	7
Dezorientacja	8	26	44	15	5	2	0
Zaburzenia uwagi	0	3	18	38	32	9	0
Brak krytycyzmu	0	1	8	28	40	16	7
Zaburzenia woli	1	2	11	44	36	6	0
Drażliwość	6	21	39	22	8	4	0
Zaabsorbowanie własn. myśl.	0	3	13	30	40	12	2
Czynne unikanie kont. społ.	0	1	14	36	14	30	5

Ocenę normatywną zacięniowano