

## METODY OCENY ISTOTNOŚCI KLINICZNEJ ZMIANY W EMPIRYCZNEJ WERYFIKACJI SKUTECZNOŚCI PSYCHOTERAPII

### CLINICAL SIGNIFICANCE METHODS IN EMPIRICAL VERIFICATION OF THE EFFECTIVENESS OF PSYCHOTHERAPY

Wydział Psychologii Uniwersytetu Warszawskiego

#### psychotherapy effectiveness clinical significance

*Artykuł przedstawia sposoby empirycznej weryfikacji skuteczności psychoterapii. Zostały w nim przede wszystkim omówione metody oceny istotności klinicznej zmian zachodzących u pacjentów biorących udział w psychoterapii. Na wstępie opisano koncepcję istotności statystycznej oraz koncepcję wielkości efektu, a w dalszej części następujące metody i koncepcje istotności klinicznej: metodę wskaźnika rzetelności zmiany i jej modyfikacje, metodę porównań normatywnych, koncepcję jakości życia, koncepcję walidacji społecznej oraz metodę Number Needed to Treat. Artykuł zawiera również dane na temat częstości korzystania z poszczególnych sposobów empirycznej weryfikacji skuteczności psychoterapii przez badaczy.*

**Summary:** This article presents the empirical ways of verifying the effectiveness of psychotherapy, mainly the methods assessing the clinical significance of change experienced by patients. At the beginning, the statistical significance and effect size are described, and in the next part the following clinical significance methods and concepts are reviewed: the reliable change index method and its modifications, the normative comparisons method, the quality of life concept, the social validation concept and the number needed to treat method. In addition, there is some data concerning the popularity of using the ways of investigating the effects of psychotherapy by researchers.

Na obecnym etapie rozwoju metodologii badań nad skutecznością psychoterapii coraz więcej badaczy próbuje na podstawie uzyskanych wyników odpowiedzieć na następujące pytania [1]:

1. Czy zaobserwowane rezultaty są prawdziwe, czy przypadkowe (związane z błędem pomiaru)?
2. Jeżeli wynik jest prawdziwy, to jak można określić jego wielkość,
3. Czy otrzymany rezultat jest na tyle duży, że można uznać go za znaczący i użyteczny?

Te trzy podejścia do analizy wyników badań nad skutecznością psychoterapii można określić odpowiednio jako: 1 — istotność statystyczną, 2 — istotność praktyczną oraz 3 — istotność kliniczną [2, 3, 4].

Niniejszy artykuł ma na celu krótkie przybliżenie ewolucji podejścia do empirycznej weryfikacji skuteczności psychoterapii oraz przedstawienie powstałych do tej pory metod badania istotności klinicznej zmiany zachodzącej u uczestników psychoterapii.

### 1. Od istotności klinicznej do... istotności klinicznej

Pierwsza strategia badań nad psychoterapią miała charakter idiograficzny. Na początku XX wieku psychoanalicy wykorzystywali studium przypadku, rozwijane następnie przez psychoterapeutów humanistycznych i fenomenologicznych [5]. Najwcześniejsze badania nad skutecznością psychoterapii opierały się głównie na opinii terapeuty o stopniu poprawy stanu pacjenta [6]. Przegląd danych zebranych w Berlińskim Instytucie Psychoanalizy w latach 20. XX wieku wskazuje na to, że psychoanalicy po zakończeniu leczenia klasyfikowali swoich pacjentów jako wyleczonych i niewyleczonych, a także takich, których stan poprawił się w mniejszym lub większym stopniu, biorąc pod uwagę takie kryteria, jak ustąpienie objawów oraz pozytywne zmiany w osobowości. Stąd wniosek, że w trakcie początków uprawiania psychoterapii terapeuci skupiali się na klinicznym wymiarze zmiany, na poprawie, którą można było rzeczywiście zaobserwować u pacjenta. Jednak wraz z rozwojem innych szkół psychoterapii pojawił się zarzut, że subiektywne zdanie psychoterapeuty nie może być miarą skuteczności jego własnych interwencji i badacze zaczęli się skupiać na naukowych metodach weryfikacji działania psychoterapii.

#### 1. 1. Istotność statystyczna

Obliczanie istotności statystycznej jest standardową procedurą podejmowaną w większości badań i od kilkudziesięciu lat stanowi ważną część metodologii badań psychologicznych. W badaniach nad skutecznością psychoterapii standardowym wymaganiam jest wykorzystanie grupy kontrolnej oraz rzetelnych i trafnych narzędzi do pomiaru badanych zmiennych z perspektywy pacjenta, znaczącej osoby z jego otoczenia lub terapeuty [5, 7]. Otrzymane w badaniach wyniki są poddawane odpowiednim analizom statystycznym [8]. Jeśli prawdopodobieństwo wystąpienia różnicy między wynikami grup jest większe niż zakładany próg błędu, przyjmuje się, że otrzymane wyniki nie są rezultatem przypadkowej zmienności dobranych prób i można je uznać za istotne statystycznie. Można zatem powiedzieć, iż analizy statystyczne i istotność statystyczna mają duże znaczenie dla oceny efektów psychoterapii, ponieważ dostarczają informacji, czy zmiany, które zaszły, są przypadkowe czy nie. Pojawiły się jednak zarzuty podające w wątpliwość użyteczność koncepcji istotności statystycznej w ocenie skuteczności leczenia pacjentów. Po pierwsze, szansę otrzymania wyników istotnych statystycznie można uczynić bardziej prawdopodobną przez zwiększenie liczebności próby, więc badając odpowiednio dużą grupę można uzyskać potwierdzenie istotnych statystycznie zależności między niemalże dowolnymi zmiennymi [8]. Po drugie, metody wnioskowania statystycznego nie określają praktycznego znaczenia efektów psychoterapii ani wielkości zaistniałej zmiany, lecz jedynie wskazują, kiedy należy odrzucić hipotezę zerową mówiącą o braku zmiany [9]. Istotna statystycznie zmiana niekoniecznie świadczy o rzeczywistych, odczuwanych przez pacjenta i jego otoczenie rezultatach psychoterapii. Po trzecie, porównania statystyczne między grupami nie pokazują lub ujawniają w bardzo małym stopniu różnorodność indy-

widualnych reakcji pacjentów na leczenie [10, 11]. Średnie grupowe nie wskazują, jaki jest procent uczestników terapii, którzy osiągnęli poprawę lub zostali wyleczeni, bądź których stan się pogorszył.

### 1. 2. Istotność praktyczna

Pewną próbą wyjścia naprzeciw zarzutom postawionym koncepcji istotności statystycznej było uzupełnienie standardowych metod statystycznych pojęciem wielkości efektu [9]. Miary wielkości efektu umożliwiają ocenę wielkości zmiany, która nastąpiła na skutek oddziaływania eksperymentalnego. Wielkość efektu może być wyrażona za pomocą wystandaryzowanej różnicy między średnimi dwóch badanych grup (np. współczynnik  $d$  Cohena) bądź siły związku między dwiema lub więcej zmiennymi (np. współczynnik  $r$  Pearsona). Wielkość efektu jest niezależna od liczebności próby, dzięki czemu można porównywać stopień ważności manipulacji eksperymentalnej w różnych badaniach oraz sporządzać metaanalizy rezultatów wielu prac. W czwartym i piątym wydaniu *The Publication Manual of the American Psychological Association* znajduje się zalecenie, aby w publikowanych pracach empirycznych obliczać i podawać wartość wielkości efektu [3]. Wiele naukowych czasopism psychologicznych wymaga od autorów raportowania tej miary w publikacjach [4].

Mimo wielu zalet wartość wielkości efektu nie pozostaje jednak zupełnie niezależna od właściwości badanych grup, jest bowiem związana z wielkością wariancji [12]. Ponadto miara ta nie uniknęła głównej słabości charakterystycznej dla istotności statystycznej. Mianowicie pojęcie to wciąż jest oderwane od praktyki, ponieważ nie pozwala na odniesienie zmierzonej zmiany do jakości funkcjonowania pacjenta w społeczeństwie czy ustąpienia objawów zaburzenia [2]. Miara ta nie wskazuje zróżnicowania wyników w grupie oraz proporcji osób w zależności od tego, jak zmienił się ich stan dzięki psychoterapii [6, 10]. Można więc uznać, że wielkość efektu jest tylko udoskonaleniem metod wnioskowania statystycznego, ale wciąż nie ma wiele wspólnego z kliniczną istotnością zmian zachodzących u pacjentów [11].

### 1. 3. Istotność kliniczna

Z powyższej krytyki wyłoniła się koncepcja istotności klinicznej. Według Kazdina istotność kliniczną „odnosi się do praktycznej lub użytecznej wartości bądź ważności efektu interwencji, tzn., czy interwencja jest przyczyną prawdziwej (autentycznej, wyraźnej, praktycznej, zauważalnej) różnicy w codziennym życiu z perspektywy klienta lub osób z jego otoczenia” [13, s. 332]. Od metod statystycznych odróżnia ją zatem to, że zamiast opierać się na ocenie zmiany na podstawie arbitralnie ustalonych wskaźników, uwzględnia ona funkcjonowanie pacjentów w kontekście społecznym oraz rzeczywiste, praktyczne znaczenie zmiany jako efektu interwencji. Pojęcie istotności klinicznej zwraca również uwagę na zmianę, jaka wystąpiła u jednostki, a nie tylko na średni poziom zmiany w całej grupie pacjentów. Powstały następujące podejścia wykorzystujące metodę istotności klinicznej [14]:

Koncepcje odwołujące się do metod statystycznych, koncentrujące się na poszukiwaniu odpowiedzi na pytania, czy zmiana osiągnięta przez jednostkę jest na tyle duża, że można

ją uznać za znaczącą (metoda wskaźnika rzetelności zmiany) [10, 11], oraz czy pacjenci, którzy przeszli leczenie, jako grupa są nie do odróżnienia od normalnie funkcjonujących osób w odniesieniu do stanu przed podjęciem psychoterapii (metoda porównań normalnych) [15].

- Koncepcja walidacji społecznej uwzględniająca opinię otoczenia społecznego o praktycznym znaczeniu leczenia i jego rezultatach [16].
- Koncepcja jakości życia oceniająca wartość interwencji psychoterapeutycznych z punktu widzenia samego pacjenta [17].
- Metoda Number Needed to Treat (NNT — liczba pacjentów wymagana do odnotowania poprawy) skupiająca się na przystępnym przedstawieniu, osobom innym niż psychoterapeuci i badacze (np. klienci, firmy ubezpieczeniowe, pracownicy służby zdrowia), jak dobrze działa psychoterapia i jakie jest prawdopodobieństwo odniesienia z niej korzyści [18].

W następnej części niniejszego artykułu metody te zostaną bardziej szczegółowo zaprezentowane.

## 2. Metody oceny istotności klinicznej efektów psychoterapii

### 2. 1. Metoda wskaźnika rzetelności zmiany

Metoda wskaźnika rzetelności zmiany opiera się na założeniu, że zmiana w psychoterapii jest istotna klinicznie, gdy pacjent pod względem badanej zmiennej przechodzi z grupy osób dysfunkcyjnych do grupy osób funkcjonujących prawidłowo [11]. Aby to zweryfikować należy, po pierwsze, ustalić rozkład zmiennej w próbie dysfunkcyjnej i próbie funkcjonalnej oraz na podstawie obu rozkładów zidentyfikować punkt odcięcia, którego przekroczenie oznacza powrót do normalnego funkcjonowania. Drugim etapem jest zbadanie rzetelności zmiany. Jeżeli rozkłady zmiennej w obu próbach zachodzą na siebie, przekroczenie punktu odcięcia niekoniecznie musi być istotne ze statystycznego punktu widzenia. Ocenę wiarygodności statystycznej umożliwia wskaźnik rzetelności zmiany (Reliable Change Index – RCI), który informuje, czy zmiana jest istotna, czy wynika jedynie z nieprecyzyjności narzędzia pomiarowego. Zmianę u konkretnego pacjenta można uznać za znaczącą tylko wtedy, gdy zostanie przekroczony próg odcięcia i jednocześnie będzie to rzetelne statystycznie.

Poniżej przedstawiono procedurę obliczania wskaźników wymaganych przez tę metodę.

a) Obliczenia punktu odcięcia dla rozkładu grupy dysfunkcyjnej i grupy funkcjonalnej można dokonać na trzy sposoby:

- Gdy niedostępne są normy dla narzędzia badawczego, identyfikację punktu odcięcia opiera się na założeniu, że poziom funkcjonowania pacjenta po zakończeniu psychoterapii powinien znaleźć się poza zakresem populacji dysfunkcyjnej, czyli 2 odchylenia standardowe (w kierunku funkcjonalności) od średniej w tej populacji.

$$a = M_1 + 2s_1$$

$M_1$  – średnia w grupie dysfunkcyjnej przed leczeniem;  $s_1$  – odchylenie standardowe w grupie dysfunkcyjnej przed leczeniem.

- Gdy dostępne są wyniki normatywne narzędzia badawczego oraz rozkłady grupy dysfunkcyjnej i grupy funkcjonalnej nie zachodzą na siebie, poziom funkcjonowania pacjenta po leczeniu powinien znaleźć się w zakresie populacji funkcjonalnej, czyli w obszarze 2 odchyłeń standardowych od średniej w tej populacji.

$$b = M_0 + 2s_1$$

$M_0$  – średnia w grupie funkcjonalnej

- Gdy dostępne są normy dla narzędzia badawczego oraz rozkłady grupy dysfunkcyjnej i grupy funkcjonalnej zachodzą na siebie, poziom funkcjonowania pacjenta powinien znaleźć się bliżej średniej w populacji funkcjonalnej niż średniej w populacji dysfunkcyjnej.

$$c = \frac{s_0 M_1 + s_1 M_0}{s_0 + s_1}$$

$s_0$  – odchylenie standardowe w grupie funkcjonalnej

- b) Gdy rozkłady grupy dysfunkcyjnej i grupy funkcjonalnej zachodzą na siebie, obliczenie wskaźnika rzetelności zmiany odbywa się według wzoru:

$$RC = \frac{x_2 - x_1}{s_{diff}}$$

$$s_{diff} = \sqrt{2(s_E)^2}$$

$$s_E = s_1 \sqrt{1 - r_{xx}}$$

$x_1$  – wynik osoby badanej przed leczeniem;  $x_2$  – wynik osoby badanej po leczeniu;  
 $r_{xx}$  – rzetelność powtórzeniowa narzędzia badawczego

- c) Na podstawie użycia kryterium przekroczenia punktu odcięcia i wartości wskaźnika rzetelności zmiany (tzn. powyżej 1,96) przeprowadzana jest klasyfikacja każdego pacjenta do jednej z grup:

- wyleczenie (recovered) — przekroczenie punktu odcięcia i wartości RCI,
- poprawa (improved) — przekroczenie wartości RCI, ale nie punktu odcięcia,
- brak zmiany (unchanged/undeterminate) — nieprzekroczenie ani punktu odcięcia, ani wartości RCI,
- pogorszenie (deteriorated) — przekroczenie wartości RCI w kierunku negatywnym.

Wyniki metody wskaźnika rzetelności zmiany Jacobsona–Truax (JT) przedstawia się jako rozkład liczbowy i procentowy pacjentów należących do poszczególnych grup.

Ta koncepcja istotności klinicznej ma jednak pewne słabości [6, 9–11, 19–25] i w związku z tym próbowano tworzyć jej modyfikacje, dokładniej szacujące znaczenie zmiany. Poniżej zostaną opisane metody alternatywne dla wskaźnika rzetelności zmiany JT.

### 2. 1. 1. Metoda Gulliksena–Lorda–Novicka (GLN)

Hsu jako pierwszy zaproponował alternatywną metodę wskaźnika rzetelności zmiany, którą oparł na pracach Lorda i Novicka oraz Linna i Slinde'a [za: 19, 26]. Zwrócił uwagę, że wskaźnik RCI Jacobsona i Truax nie bierze pod uwagę zjawiska regresji do średniej, które polega na tym, że skrajne wyniki z powodu niedoskonałej rzetelności narzędzia pomiarowego będą się stawać mniej skrajne w wyniku procedury powtórzonych pomiarów. W związku z tym wynik uzyskany po psychoterapii przez każdą osobę w badanej grupie będzie bliższy średniej grupowej niż wynik sprzed leczenia. To z kolei może przyczynić się do błędnej klasyfikacji pacjentów do grup w zależności od zmiany doświadczonej na skutek psychoterapii. Metoda GLN stara się temu zapobiec przez uwzględnienie średniej badanej zmiennej w populacji, do której uzyskane wyniki się zbliżają, np. dzięki wykorzystaniu średniej oszacowanej dla grupy, z której pochodzą uczestnicy badania, bądź wyników osób badanych przed rozpoczęciem psychoterapii. Wzór wskaźnika RCI według metody GLN ma następującą postać:

$$\frac{[X_{post} - M_{pop}] - r_{xx}[X_{pre} - M_{pop}]}{S_{pop}\sqrt{(1 - r_{xx}^2)}}$$

$X_{post}$  – wynik surowy osoby badanej po leczeniu;  $X_{pre}$  – wynik surowy osoby badanej przed leczeniem;  $M_{pop}$  – hipotetyczna średnia w populacji, do której zbliżają się wyniki;  $S_{pop}$  – odchylenie standardowe populacji, do którego zbliżają się wyniki;  $r_{xx}$  – rzetelność powtórzeniowa narzędzia badawczego

### 2. 1. 2. Metoda Edwardsa–Nunnally'ego (EN)

Autorem metody EN jest Speer, który do jej stworzenia wykorzystał prace Edwardsa, Yarvisa, Muellera, Zingale i Wagmana oraz Nunnally'ego [24]. Podobnie jak w metodzie GLN, sposób obliczania wskaźnika RCI wg EN zmniejsza wpływ regresji do średniej na ocenę istotności zmian zachodzących podczas leczenia. Za pomocą metody EN określa się, czy zmiana jest rzetelna przez odniesienie indywidualnego wyniku otrzymanego po zakończeniu psychoterapii do przedziału ufności ( $\pm 2$  błędy standardowe pomiaru) ustalonego dla oszacowanego prawdziwego wyniku pacjenta sprzed leczenia. Jeżeli wynik osoby badanej po zakończeniu psychoterapii znajdzie się poza przedziałem ufności, to ten rezultat może zostać uznany za istotnie różny od wyniku przed uczestnictwem w psychoterapii. Wskaźnik RCI według tej metody oblicza się za pomocą wzoru:

$$[r_{xx}(X_{pre} - M_{pre}) + M_{pre}] \pm 2S_{pre}\sqrt{1 - r_{xx}}$$

$r_{xx}$  – rzetelność powtórzeniowa narzędzia badawczego;  $X_{pre}$  – wynik surowy osoby badanej przed leczeniem;  $M_{pre}$  – średnia grupy przed leczeniem;  $S_{pre}$  – odchylenie standardowe grupy przed leczeniem

### 2. 1. 3. Metoda Hagemana–Arrindella (HA)

Hageman i Arrindell [27] opierając się na pracy Cronbacha i Glesera zmodyfikowali metodę Jacobsona i Truax w jeszcze większym stopniu niż Hsu i Speer. Metoda HA umożliwia analizę zmiany pod względem jej istotności klinicznej zarówno na poziomie indywidualnym, jak i na poziomie grupowym (obliczanie procentu uczestników psychoterapii, którzy doświadczyli rzetelnej zmiany i przekroczyli ustalony punkt odcięcia). Nowy wskaźnik rzetelności zmiany,  $RC_{INDIV}$ , według autorów tej metody, charakteryzuje się większą precyzją szacowania znaczenia zmiany niż wskaźnik JT, gdyż bierze pod uwagę rzetelność obu pomiarów: przed leczeniem i po leczeniu. Ponadto metoda HA zawiera modyfikację wyznaczania punktu odcięcia ( $CS_{INDIV}$ ). Po pierwsze, uwzględnia poprawki dla regresji do średniej, podobnie jak w przypadku obliczania  $RC_{INDIV}$ , a po drugie — dla każdej osoby jest szacowany indywidualny punkt odcięcia.

Procedura stosowania metody HA jest złożona i można ją odnaleźć w artykule Hagemana i Arrindella [27].

### 2. 1. 4. Metoda hierarchicznego modelowania liniowego (Hierarchical Linear Modelling – HLM)

Metoda HLM została rozwinięta przez Speera i Greenbauma; opiera się na modelu wzrostu krzywej [28]. Bierze pod uwagę więcej niż dwa pomiary zmiennej u pacjenta, co zdaniem autorów pozwala na dokładniejszą ocenę zmiany zachodzącej między wynikami uzyskanymi przed leczeniem i po leczeniu. Metoda ta różni się od poprzednich tym, że wykorzystuje oszacowania Bayesowskie, które są ważonymi wartościami łączącymi dane pochodzące od jednostki oraz od całej próby, a także tym, że ma elastyczne wymagania wobec danych, tzn. dopuszcza braki danych w poszczególnych pomiarach. Metoda HLM polega na obliczeniu indywidualnych parametrów wzrostu krzywej dla pacjenta, np. nachylenia linii czy wartości współczynników zmiany. Ostateczny wynik uzyskuje się według wzoru:

$$\frac{B^*}{\sqrt{V^*}}$$

$B^*$  — empiryczne oszacowanie Bayesa nachylenia krzywej;  $\sqrt{V^*}$  — odchylenie standardowe empirycznego oszacowania Bayesa

W środowisku naukowym trwają dyskusje, która metoda określania rzetelności zmiany jest najlepsza i najbardziej precyzyjna oraz w największym stopniu zgodna z koncepcją istotności klinicznej [10, 19, 21, 24–34]. Badania nie wskazują jednak jednoznacznych ani znaczących różnic między wynikami uzyskiwanymi przez pięć opisanych metod. Autorzy każdej z nich starają się udowodnić wyższość własnej koncepcji nad pozostałymi. Niektórzy badacze są zdania, że skoro metody nie różnią się wzajemnie w znaczącym stopniu, zalecane jest wykorzystywanie oryginalnej metody Jacobsona i Truax ze względu na jej popularność i najmniej skomplikowaną procedurę stosowania oraz sugerują skupienie się na dokładnym zbadaniu powstałych dotychczas metod, zamiast tworzeniu nowych [3, 19, 35].

## 2. 2. Metoda porównań normatywnych

Według metody porównań normatywnych skutkiem istotnej klinicznie zmiany jest niemożność odróżnienia grupy pacjentów, którzy ukończyli terapię, od niezaburzonej grupy odniesienia dobrze funkcjonującej pod względem badanych zmiennych [15]. Aby to zbadać, wykorzystuje się testy ekwiwalencji — zaadaptowane z badań nad działaniem leków — których zadaniem jest wykrycie, czy wyniki dwóch grup są na tyle zbliżone, aby móc uznać je za równoważne. Wobec tego testy ekwiwalencji pełnią przeciwną rolę do testów istotności statystycznej określających, czy wyniki dwóch grup różnią się istotnie. Dopuszczalna odległość między rozkładami wyników dwóch grup, wyrażona najczęściej w jednostkach odchylenia standardowego, wyznaczająca ich ekwiwalencję, może być określona na podstawie analizy norm populacyjnych lub wyników uzyskiwanych w wykorzystywanych narzędziach pomiarowych. Gdy testy ekwiwalencji są stosowane w połączeniu z tradycyjnym testowaniem hipotez, pozwala to na jeszcze bardziej pewne wnioskowanie dotyczące równoważności grup.

Procedura wykonania porównań normatywnych składa się z następujących etapów:

- a) Ustalenie zakresu (poniżej i powyżej różnicy między średnią grupy normatywnej i średnią grupy klinicznej po leczeniu), w jakim muszą się znaleźć średnie wyniki obu grup, aby można je było uznać za ekwiwalentne ( $\delta_1$  i  $\delta_2$ ).
- b) Przeprowadzenie testów  $t$  lub  $z$  (w zależności od liczebności próby i wiedzy o odchyleniu standardowym w populacji) ekwiwalencji średnich według wzorów:

$$C.E.t(df) = \left( \frac{M_N - M_C - \delta_1}{SE_{N-C}} \right)$$

$$C.E.t(df) = \left( \frac{M_N - M_C - \delta_2}{SE_{N-C}} \right)$$

$$C.E.z_{un} = \left( \frac{M_N - M_C - \delta_1}{SE_{N-C}} \right)$$

$$C.E.z_{un} = \left( \frac{M_N - M_C - \delta_2}{SE_{N-C}} \right)$$

$$SE_{N-C} = \left\{ \left[ \frac{(n_N - 1)SD_N^2 + (n_C - 1)SD_C^2}{n_N + n_C - 2} \right] \left[ \frac{1}{n_N} + \frac{1}{n_C} \right] \right\}^{1/2}$$

$M_N$  – średnia grupy normatywnej;  $M_C$  – średnia grupy klinicznej;  $SE_{N-C}$  – błąd standardowy różnicy między średnimi grupy normatywnej i klinicznej;  $n_N$  – liczebność grupy normatywnej;  $n_C$  – liczebność grupy klinicznej;  $SD_N$  – odchylenie standardowe w grupie normatywnej;  $SD_C$  – odchylenie standardowe w grupie klinicznej;  $C.E.$  – kliniczna ekwiwalencja



Przeprowadza się tylko jeden jednostronny test ekwiwalencji o bardziej rygorystycznej wartości  $\delta$ .

c) Wykonanie tradycyjnego testu istotności różnic między średnimi  $t$  lub  $z$  według wzoru:

$$t(df) = \left( \frac{M_N - M_C}{SE_{N-C}} \right)$$

$$z_{un} = \left( \frac{M_N - M_C}{SE_{N-C}} \right)$$

d) Klasyfikacja i interpretacja wyników testów tradycyjnych i testów ekwiwalencji:

		Tradycyjny test statystyczny	
		Wynik istotny	Wynik nieistotny
Test ekwiwalencji klinicznej	Wynik istotny	<i>Statystycznie różne, klinicznie ekwiwalentne</i>	<i>Klinicznie ekwiwalentne</i>
	Wynik nieistotny	<i>Różne (nieekwiwalentne klinicznie)</i>	<i>Wyniki niejednoznaczne (wymagana większa moc – większa próba/mniejsza wariancja wyników)</i>

Cribbie i Arpin-Cribbie [36] wprowadzili wiele modyfikacji do standardowej metody porównań normatywnych w celu jej ulepszenia. Po pierwsze, sugerują wykonywanie dwóch testów ekwiwalencji, gdyż od różnicy między średnimi zależy, który z nich jest bardziej rygorystyczny. Po drugie, zalecają stosowanie kilku przedziałów ekwiwalencji (definitywna ekwiwalencja, możliwa ekwiwalencja, potencjalna ekwiwalencja). Po trzecie, podają wzór na test ekwiwalencji Schuirmana–Welcha, którego należy używać, gdy porównywane grupy różnią się pod względem liczebności i wariancji, ponieważ wyniki zwykłego testu ekwiwalencji mogą być w takiej sytuacji zniekształcone. Po czwarte, podważają sens przeprowadzania tradycyjnych testów różnic na średnich wynikach dwóch grup po zakończeniu psychoterapii i sugerują wykonanie tylko testów ekwiwalencji, gdyż to one tak naprawdę weryfikują hipotezę o równoważności średnich.

### 2. 3. Koncepcja jakości życia

Metoda istotności klinicznej zmian zachodzących w psychoterapii pod względem jakości życia jest uzupełnieniem opisanych wcześniej koncepcji skupiających się na zweryfikowaniu rzetelności zmiany i jej praktycznego znaczenia za pomocą metod statystycznych. U podstaw tej koncepcji leży spojrzenie na zdrowie z perspektywy pozytywnej [37], jako „pełnego dobrostanu fizycznego, psychicznego i społecznego, a nie tylko nieobecności choroby” [za: 17]. Dlatego też w centrum zainteresowania koncepcji jakości życia, w przeciwieństwie do metod Jacobsona i Truax czy Kendalla i współpracowników,

nie leży ustąpienie objawów zaburzeń dzięki leczeniu, lecz wpływ uczestnictwa w psychoterapii na poziom dobrostanu i zadowolenia z codziennego życia z punktu widzenia pacjenta [17].

Po pierwsze, zmianę w jakości życia można zweryfikować w sposób subiektywny, kiedy to pacjent – „konsument” usługi, jaką jest psychoterapia, ocenia jej efekty pod kątem tego, czy zaszła pożądana przez niego zmiana i poprawiła jego zadowolenie z życia [16]. Drugi sposób, wymieniany przez Sechresta, McKnighta i McKnight, ma charakter bardziej obiektywny i funkcjonalny, gdyż opiera się na możliwych do zaobserwowania zmianach w funkcjonowaniu społecznym pacjenta jako lepszej podstawie oceny efektów leczenia niż pomiar zmiany za pomocą narzędzi badawczych stworzonych na podstawie niesprecyzowanych konstruktów teoretycznych [38]. Rzeczywistą zmianą w funkcjonowaniu pacjenta może być podejmowanie zachowań, których nie potrafił przejawiać przed psychoterapią, ale jest w stanie podejmować po zakończeniu leczenia, np. zawarcie większej liczby znajomości w ciągu miesiąca przez osobę z fobią społeczną.

Pojęcie jakości życia nie jest jednoznaczne i istnieje wiele jego wersji [17, 37]. Z tego względu powstało również bardzo dużo narzędzi do pomiaru tego konstruktów. Gill i Feinstein raportują na podstawie przeanalizowania 75 badań, że zostało w nich użytych 159 miar jakości życia [za: 17]. Narzędzia do oceny tego pojęcia można podzielić na dwie grupy: oparte na teorii ogólnej jakości życia i oparte na jakości życia związanej ze zdrowiem. Teoria ogólnej jakości życia dotyczy aspiracji, celów i radzenia sobie jednostki w odniesieniu do zarówno wewnętrznych standardów, jak i porównania z innymi ludźmi. Z kolei koncepcja jakości życia związanej ze zdrowiem koncentruje się na obciążeniu związanym z chorobą: jej objawami, osłabieniem funkcjonowania oraz leczeniem.

#### 2. 4. Koncepcja walidacji społecznej

Koncepcja walidacji społecznej skupia się na istotności klinicznej zmian zachodzących w trakcie psychoterapii z perspektywy społeczeństwa, biorąc pod uwagę fakt, że społeczne normy i kryteria oceny skuteczności psychoterapii mogą się różnić od poglądów profesjonalistów [16]. Podejście to postuluje przystępne dla zwykłych ludzi sposoby określania, czy interwencje psychologiczne są w stanie pomóc w rozwiązywaniu problemów, oraz przyznaje społeczeństwu – jako pośredniemu „konsumentowi” psychoterapii – ważną rolę w nauce.

Walidacja społeczna może się odbywać na trzech płaszczyznach [39]:

- ocena znaczenia celów psychoterapii,
- ocena akceptacji interwencji stosowanych w trakcie leczenia,
- ocena znaczenia zmiany zachowania pacjenta jako efektu psychoterapii.

W związku z tym weryfikacja skuteczności leczenia z punktu widzenia społeczeństwa składa się z kilku etapów.

Po pierwsze, psychoterapeuci mogą się konsultować z otoczeniem pacjenta, aby dowiedzieć się, nad jakimi trudnościami powinni z nim pracować oraz jakie cele psychoterapii proponują najbliższe mu osoby [16]. Jest to cenne źródło informacji o tym, jakie zachowania otoczenie pacjenta uważa za pozytywne, a jakie za negatywne, i może

posłużyć do wyboru wymiarów, na podstawie których będzie oceniane funkcjonowanie uczestnika psychoterapii. Ponadto można również zbadać różne zachowania występujące w otoczeniu pacjenta, aby osądzić, jaki poziom tych zachowań jest na tyle skrajny, że wymaga leczenia [za: 22].

Po drugie, stopień akceptacji interwencji psychoterapeutycznych można poznać, prezentując studentom, potencjalnym klientom lub pracownikom służby zdrowia, hipotetyczne sytuacje opisujące problem pacjenta i możliwe procedury leczenia oraz poproszenie respondentów o ocenę tego planu [16]. Inną drogą jest poproszenie pacjentów po udziale w psychoterapii o ocenę konkretnych aspektów leczenia (np. stosowane interwencje, osoba terapeuty) za pomocą różnych narzędzi do pomiaru satysfakcji.

Po trzecie, istotność kliniczną efektów psychoterapii można szacować na dwa sposoby. Jeden z nich to metoda porównań społecznych polegająca na odniesieniu zachowań pacjenta po leczeniu do zachowań grupy osób z jego otoczenia funkcjonujących prawidłowo pod względem badanych wymiarów [za: 22] (ten sposób jest tożsamy z metodą porównań normatywnych, opisaną w części 2.2.). Drugi opiera się na subiektywnej ocenie funkcjonowania pacjenta — czy zmiany jakościowe są zauważalne — z perspektywy jego samego, bliskich osób utrzymujących z nim częsty kontakt lub profesjonalistów. Zmianę określa się jako istotną klinicznie, jeśli jeden z tych dwóch sposobów wskazuje na to, że w istotnych dla życia pacjenta wymiarach znacznie zmniejszyła się częstość zachowań niepożądanych na rzecz zachowań pożądaných, w porównaniu ze stanem sprzed psychoterapii. Foster i Mash zwracają uwagę na to, że walidacja społeczna nie jest konstruktem dychotomicznym, lecz ciągłym [16]. Interwencji psychoterapeutycznych nie dzieli się na mające i niemające znaczenia społecznego — ich społeczną wartość ocenia się na wielowymiarowym kontinuum.

## 2. 5. Number Needed to Treat (NNT)

Koncepcja NNT, której autorami są Laupacis, Sackett i Roberts, służy do przedstawienia statystycznej oraz klinicznej istotności interwencji [18]. **Ważnym celem tej metody jest intuicyjność interpretacji wyników**, aby osoby niezajmujące się profesjonalnie psychoterapią bądź badaniem jej skuteczności były w stanie wyciągnąć zrozumiałe wnioski dotyczące jakości leczenia. Wskaźnik NNT wyraża liczbę osób, które należy poddać konkretnemu rodzajowi psychoterapii, zanim jedna osoba odniesie z niej korzyści. Podanie wielkości NNT wymaga obliczenia wartości wskaźnika ARR (Absolute Risk Reduction), czyli absolutnej redukcji ryzyka, będącej różnicą między proporcją badanych w grupie kontrolnej a proporcją osób biorących udział w psychoterapii, których stan nie uległ poprawie bądź się pogorszył [40]. NNT jest odwrotnością AAR i jego wartość otrzymuje się w wyniku podzielenia 1 przez AAR, co wyraża poniższy wzór [41]:

$$NNT = \frac{1}{ARR}$$

$$ARR = D_c - D_i$$

$D_c$  — stosunek osób w grupie kontrolnej, których stan nie uległ poprawie lub pogorszył się, do wszystkich osób w grupie kontrolnej;  $D_i$  — stosunek osób poddanych interwencji, których stan nie uległ poprawie lub pogorszył się, do wszystkich osób poddanych interwencji.

Wartość NNT bliska 1 wskazuje na to, że efekty psychoterapii były korzystne dla prawie wszystkich jej uczestników, natomiast wartość około 2—3 sugeruje, że leczenie ogólnie było skuteczne [40]. Obecnie nie istnieje granica wartości wskaźnika NNT, od której interwencję można ocenić jako nieskuteczną [18]. Leung zaleca raportowanie ARR i NNT z przedziałami ufności, w których znajdują się prawdziwe wartości wskaźników przy założonym poziomie istotności statystycznej [42].

NNT umożliwia łatwe porównanie skuteczności różnych rodzajów leczenia tych samych zaburzeń czy objawów chorobowych [43]. Dla przykładu, gdy dla jednej interwencji wskaźnik NNT wynosi 5, a dla drugiej 10, można wyciągnąć wniosek, że pierwszą metodą charakteryzuje większe prawdopodobieństwo otrzymania pozytywnych efektów psychoterapii. Jednak, aby móc porównać wartości wskaźników NNT, uzyskane w różnych badaniach, należy mieć pewność, że została wykorzystana taka sama miara efektów leczenia, reakcja na interwencje została zdefiniowana w taki sam sposób oraz grupa kontrolna i grupa leczonych pacjentów były równoważne pod względem zmiennych społeczno-demograficznych [44].

### 3. Jak często badacze stosują metody oceny istotności klinicznej efektów psychoterapii?

Faulkner, Fidler i Cumming [45] dokonali przeglądu randomizowanych badań kontrolnych nad skutecznością psychoterapii, opublikowanych w popularnych czasopismach psychologicznych i psychiatrycznych, pod kątem rodzaju raportowanych statystyk [45]. Praca obejmuje okres od 1999 do 2003 roku, a więc czas, gdy określanie istotności klinicznej zmian zachodzących podczas psychoterapii stawało się coraz bardziej popularne w środowisku naukowym, a w piątym wydaniu APA *Publication Manual* został położony nacisk na podawanie wielkości efektu. Z przeglądu wynika, że metody wnioskowania statystycznego są nieodłącznym elementem analizy wyników badań, zarówno w psychologii (99%), jak i w psychiatrii (95%). Jeżeli chodzi o miary wielkości efektu, to publikacje psychiatryczne (64%) przeważają nad psychologicznymi (43%) w ich raportowaniu, jednak więcej było artykułów psychologicznych (31%) niż psychiatrycznych (12%) zawierających ich interpretację. Największa różnica między psychologią a psychiatrią pojawiła się w przypadku wykorzystania metod istotności klinicznej — w 26% publikacji z psychologii autorzy dokonali weryfikacji istotności klinicznej zmian w psychoterapii, a takich publikacji psychiatrycznych był tylko 1%. Shearer-Underhill i Marker zbadali natomiast częstość używania metody NNT do oceny istotności klinicznej zmian zachodzących w wyniku leczenia [18]. Uwzględnili artykuły, które ukazały się w czasopiśmie *Journal of Consulting and Clinical Psychology* między rokiem 2000 a 2008. Na 100 wybranych losowo publikacji tylko 4 zawierały odwołanie do tej metody oraz wartość wskaźnika NNT.

#### Podsumowanie

Różnorodność opisanych w niniejszym artykule metod oceny istotności klinicznej zmian zachodzących w psychoterapii świadczy o złożoności tego zagadnienia. Można

zauważyć, że omawiane koncepcje koncentrują się na różnych etapach weryfikacji skuteczności leczenia. Metody wskaźnika rzetelności zmiany czy metoda porównań normatywnych, które za Jaworską [9] można zakwalifikować do nurtu obiektywnego istotności klinicznej, wskazują statystyczne procedury pomagające badaczowi sprawdzić, czy dwa wyniki (wynik pacjenta przed psychoterapią vs po psychoterapii bądź wynik grupy pacjentów po psychoterapii vs wynik osób zdrowych) są w wystarczającym stopniu odmienne bądź równoważne, aby móc tę różnicę uznać za istotną klinicznie. Te koncepcje nie zawierają natomiast zaleceń co do wymiarów, na podstawie których powinny być mierzone efekty psychoterapii, choć w ich przypadku istotność kliniczną zmiany ocenia się głównie w odniesieniu do wymiarów negatywnych dotyczących obecności objawów zaburzeń. Z kolei koncepcje jakości życia i walidacji społecznej, zaliczane przez Jaworską do nurtu subiektywnego [9], proponują wymiary oceny psychoterapii oparte na pozytywnych zmianach w funkcjonowaniu i samopoczuciu pacjenta, ale nie oferują gotowych procedur, których wykonanie dałoby wynik sugerujący, czy zmiana zachodząca dzięki psychoterapii jest praktyczna i zauważalna w codziennym życiu. Wreszcie metodę NNT, na podstawie powyższych kryteriów, można uznać za obiektywną metodę istotności klinicznej.

Różnice występujące między nurtem obiektywnym a subiektywnym sprawiają, że koncepcje istotności klinicznej uzupełniają się. Stworzone dotychczas koncepcje łączą trzy perspektywy oceny skuteczności psychoterapii [46]:

- perspektywę społeczeństwa (koncepcja walidacji społecznej),
- perspektywę klienta (koncepcja jakości życia) oraz
- perspektywę specjalistów od zdrowia psychicznego (koncepcje wskaźnika rzetelności zmiany, porównań normatywnych i NNT).

Ciągły rozwój tych koncepcji oraz powstawanie nowych świadczy o zainteresowaniu badaczy doskonaleniem pomiaru skuteczności leczenia oraz o zapotrzebowaniu osób prowadzących psychoterapię i pacjentów na informacje o praktycznym znaczeniu zmiany jako efektu interwencji. Mimo to metody istotności klinicznej w analizie wyników badań są stosowane przez niewielką część środowiska naukowego, a w Polsce jest to zagadnienie nadal mało znane w psychologii.

### Piśmiennictwo

1. Kirk RE. Promoting good statistical practices: some suggestions. *Educ. Psychol. Measur.* 2000; 61: 213–218.
2. Conner BT. When is the difference significant? Estimates of meaningfulness in clinical research. *Clin. Psychol. Sc. Pract.* 2010; 17: 52–57.
3. Pinteá S. The relevance of results in clinical research: statistical, practical and clinical significance. *J. Cogn. Behav. Psychother.* 2010; 10: 101–114.
4. Thompson B. ‘Statistical,’ ‘practical,’ and ‘clinical,’: how many kinds of significance do counselors need to consider? *J. Couns. Develop.* 2002; 80: 64–71.
5. Jakubowska U. Modele badań nad psychoterapią. W: Grzesiuk L, red. *Psychoterapia. Badania i szkolenie*. Warszawa: ENETEIA Wydawnictwo Psychologii i Kultury; 2005, s. 21–43.
6. Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: history, application, and current practice. *Clin. Psychol. Rev.* 2001; 21: 421–446.

7. Rakowska JM. Skuteczność psychoterapii. Warszawa: SCHOLAR; 2005.
8. Brzeziński J. Metodologia badań psychologicznych. Warszawa: Wydawnictwo PWN; 2004.
9. Jaworska A. Istotność zmiany w koncepcjach pomiaru skuteczności psychoterapii. W: Brzeziński J, Siuta J, red. Metodologiczne i statystyczne problemy psychologii. Poznań: Wydawnictwo Zysk i S-ka; 2006, s. 119–156.
10. Jacobson NS, Roberts LJ, Berns SB, McGlinchey JB. Methods for defining and determining the clinical significance of treatment effects: description, application, and alternatives. *J. Cons. Clin. Psychol.* 1999; 67: 300–307.
11. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J. Cons. Clin. Psychol.* 1991; 59, 12–19.
12. Follette WC, Callaghan GM. The evolution of clinical significance. *Clin. Psychol. Sc. Pract.* 2001; 8: 431–435.
13. Kazdin AE. The meanings and measurement of clinical significance. *J. Cons. Clin. Psychol.* 1999; 67: 332–339.
14. Kendall PC. Clinical significance. *J. Cons. Clin. Psychol.* 1999; 67: 283–284.
15. Kendall PC, Marrs-Garcia A, Nath SR, Sheldrick RC. Normative comparisons for the evaluation of clinical significance. *J. Cons. Clin. Psychol.* 1999; 67: 285–299.
16. Foster SL, Mash EJ. Assessing social validity in clinical treatment research: issues and procedures. *J. Cons. Clin. Psychol.* 1999; 67: 308–319.
17. Gladis MM, Gosch EA, Dishuk NM, Crits-Christoph P. Quality of life: expanding the scope of clinical significance. *J. Cons. Clin. Psychol.* 1999; 67: 320–331.
18. Shearer-Underhill C, Marker C. The use of the number needed to treat (NNT) in randomized clinical trials in psychological treatment. *Clin. Psychol. Sc. Pract.* 2010; 17: 41–47.
19. Bauer S, Lambert MJ, Nielsen SL. Clinical significance methods: a comparison of statistical techniques. *J. Person. Assess.* 2004; 82: 1, 60–70.
20. Campbell A. Clinical significance in real world settings. *Austr. NZ J. Fam. Ther.* 2008; 29: 107–110.
21. Hageman WJMM, Arrindell WA. Establishing clinically significant change: increment of precision and the distinction between individual and group level of analysis. *Beh. Res. Ther.* 1999; 1169–1193.
22. Hansen NB, Lambert MJ. Clinical significance: an overview of methods. *J. Mental Health* 1996; 5: 17–24.
23. Lunnen KM, Ogles BM. A multiperspective, multivariable evaluation of reliable change. *J. Cons. Clin. Psychol.* 1998; 66: 400–410.
24. Speer DC. Clinically significant change: Jacobson and Truax (1991) Revisited. *J. Cons. Clin. Psychol.* 1992; 60: 402–408.
25. Wise EA. Methods for analyzing psychotherapy outcomes: a review of clinical significance, reliable change, and recommendations for future directions. *J. Person. Assess.* 2004; 82: 50–59.
26. Hsu LM. Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *J. Cons. Clin. Psychol.* 1995; 63: 141–144.
27. Hageman WJMM, Arrindell WA. Clinically significant and practical! Enhancing precision does make a difference. Reply to McGlinchey and Jacobson, Hsu, and Speer. *Beh. Res. Ther.* 1999; 3: 1219–1233.
28. Speer DC, Greenbaum PE. Five methods for computing significant individual client change and improvement rates: support for an individual growth curve approach. *J. Cons. Clin. Psychol.* 1995; 63: 1044–1048.

29. Atkins DC, Bedics JD, McGlinchey JB, Beauchaine TP. Assessing clinical significance: does it matter which method we use? *J. Cons. Clin. Psychol.* 2005; 73: 982–989.
30. Hsu LM. A comparison of three methods of identifying reliable and clinically significant client changes: commentary on Hageman and Arrindell. *Behav. Res. Ther.* 1999; 37: 1195–1202.
31. Hsu LM. Caveats concerning comparisons of change rates obtained with five methods of identifying significant client changes: comment on Speer and Greenbaum (1995). *J. Cons. Clin. Psychol.* 1999; 67: 594–598.
32. Lambert MJ, Ogles BM. Using clinical significance in psychotherapy outcome research: the need for a common procedure and validity data. *Psychother. Res.* 2009; 19: 493–501.
33. McGlinchey JB, Atkins DC, Jacobson NS. Clinical significance methods: which one to use and how useful are they? *Behav. Ther.* 2002; 33: 529–550.
34. Speer DC. What is the role of two-wave designs in clinical research? Comment on Hageman and Arrindell. *Beh. Res. Ther.* 1999; 37: 1203–1210.
35. McGlinchey JB, Jacobson NS. Clinically significant but impractical? A response to Hageman and Arrindell. *Beh. Res. Ther.* 1999; 3: 1211–1217.
36. Cribbie RA, Arpin-Cribbie CA. Evaluating clinical significance through equivalence testing: extending the normative comparisons approach. *Psychother. Res.* 2009; 19: 677–686.
37. Frisch MB. Quality of life therapy and assessment in health care. *Clin. Psychol. Sc. Pract.* 1998; 5: 19–40.
38. Sechrest L, McKnight P, McKnight K. Calibration of measures for psychotherapy outcome studies. *Am. Psychol.* 1996; 51: 1065–1071.
39. Wolf MM. Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *J. App. Behav. Anal.* 1978; 11: 203–214.
40. Compton SN, March JS, Brent D, Albano AM, Weersing A, Curry J. Cognitive-behavioral psychotherapy for anxiety and depressive disorders in children and adolescents: an evidence-based medicine review. *J. Am. Acad. Child Adolesc. Psych.* 2004; 43: 930–959.
41. Christensen PM, Kristiansen IS. Number-Needed-to-Treat (NNT). – needs treatment with care. *Bas. Clin. Pharmacol. Toxicol.* 2006; 99: 12–16.
42. Leung WC. Balancing statistical and clinical significance in evaluating treatment effects. *Postgrad. Med. J.* 2001; 77: 201–204.
43. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann. Int. Med.* 1997; 126: 712–720.
44. Marrs-Garcia A. What does „work” mean? Reopening the debate about clinical significance. *Clin. Psychol. Sc. Pract.* 2010; 17: 48–51.
45. Faulkner C, Fidler F, Cumming G. The value of RCT evidence depends on the quality of statistical analysis. *Beh. Res. Ther.* 2008; 46: 270–281.
46. Strupp HH, Hadley SW. A tripartite model of mental health and therapeutic outcomes: with special reference to negative effects in psychotherapy. *Am. Psychol.* 1977; 32: 187–196.

Adres: akemi@tlen.pl