

W jaki sposób można poprawić jakość bez mierzenia wyników?

Mark Zimmerman, MD, Diane Young, PhD, Iwona Chelminski, PhD, Kirsty Dalrymple, PhD

- Wysiłki zmierzające do poprawy jakości w psychiatrii będą wymagać zmiany paradygmatu w kierunku uwzględniania wymiernych metod oceny wyników leczenia w praktyce klinicznej.
- Sześć potencjalnych przeszkód w dążeniu do leczenia opartego na pomiarach wyników to: akceptacja pacjenta, akceptacja lekarza, kliniczna przydatność pomiarów, koszty, heterogenność diagnostyczna i gromadzenie danych.
- Aby rozwiązać problem heterogenności, wprowadzono pojęcie psychiatrycznych objawów życiowych. Zaleca się rutynową ocenę objawów depresji i lęku, analogicznie do rutynowej oceny temperatury ciała i ciśnienia tętniczego na oddziałach ogólnych.
- Opisano sieciowy system zarządzania oceną wyników. Umożliwia on gromadzenie informacji na temat wyników dla różnych grup chorych. Opisano rzetelność i trafność sieciowego zarządzania.
- Opisano kilka sytuacji klinicznych ilustrujących, w jaki sposób opieka oparta na pomiarach może poprawić wynik leczenia.

Streszczenie

Warunkiem poprawy jakości jest pomiar wyników. Psychiatria jest jedyną specjalnością medyczną, w której ilościowe parametry oceny wyników leczenia nie są standardem. Pomiar wyników w rutynowej praktyce klinicznej napotyka kilka potencjalnych przeszkód, jednak ostatnio opracowano metody oceny mające na celu ich pokonanie. Jedną z takich przeszkód jest ustalenie wyników w niejednorodnych diagnostycznie grupach chorych. Przedstawiono koncepcję psychiatrycznych objawów życiowych. U przeważającej większości dużych, niejednorodnych diagnostycznie grup pacjentów psychiatrycznych korzystających z opieki ambulatoryjnej wyjściowo występowały istotne klinicznie objawy lęku lub depresji. Przemawia to za zasadnością rutynowej oceny objawów lęku i depresji w praktyce klinicznej, ponieważ podczas pierwszej wizyty u prawie wszystkich chorych stwierdza się objawy tych zaburzeń. Nawet w przypadku pacjentów bez objawów depresji i lęku, można powiedzieć, że ocena tych objawów jest istotna, analogiczna do przeprowadzanej w opiece podstawowej ocenie niektórych parametrów fizjologicznych – takich jak ciśnienie tętnicze i temperatura ciała – niezależnie od przyczyny wizyty. Inną przeszkodą w wykorzystywaniu parametrów do oceny jakości są trudności w gromadzeniu danych od dużej liczby pacjentów. Opisano internetowy system oceny wyników, który umożliwia obliczenie częstości reakcji na leczenie i remisji w grupie pacjentów. Obecnie dostępne są narzędzia i techniki, za pomocą których można przezwyciężyć trudności we wprowadzaniu do opieki nad pacjentami podejścia opartego na mierzeniu parametrów, co, miejmy nadzieję, przyspieszy wprowadzenie pomiarów do rutynowej praktyki klinicznej i umożliwi przetestowanie w ekonomiczny sposób prób poprawy jakości leczenia.

Wprowadzenie

Każdego roku *U.S. News and World Report*¹ publikuje listę najlepszych programów w psychiatrii (i innych specjalnościach medycznych). W towarzyszącym artykule zwraca uwagę brak omówienia danych opisujących częstość reakcji na leczenie. Czy pacjenci z zaburzeniami psychicznymi leczeni w ramach programów lepiej ocenianych częściej osiągają lepsze wyniki leczenia niż chorzy leczeni w ramach programów ocenianych gorzej? Czy chorzy leczeni w ramach programów, które nie są wymienione na tej liście, osiągają jeszcze gorsze rezultaty? Być może w niektórych specjalnościach medycznych są

dane wykazujące to, że w niektórych ośrodkach medycznych częściej niż w innych osiąganymi są dobre wyniki leczenia, jednak autorzy tego artykułu nie znają takich informacji w odniesieniu do psychiatrii.

Rozważając szerzej problem wyników leczenia, w dziedzinie psychiatrii dopiero zaczynamy zadawać podstawowe pytania dotyczące skuteczności dostępnych obecnie metod leczenia w rzeczywistej praktyce medycznej. Jak dobrze działają? U których chorych są najskuteczniejsze, a u których nieskuteczne? Jak wielu pacjentów jest objętych leczeniem opartym na dowodach, a jak wielu nie? Czy stosowanie leczenia opartego

Dr Zimmerman, associated professor of Psychiatry and Human Behavior, Brown University Medical School, Providence, Rhode Island. Dr Young, staff psychologist, Department of Psychiatry, Rhode Island Hospital, assistant professor of Research, Department of Psychiatry, Alpert Medical School, Brown University. D Chelminski, clinical assistant professor, Department of Psychiatry, Rhode Island Hospital. Dr Dalrymple, staff psychologist, Department of Psychiatry, Rhode Island Hospital; assistant professor of Research, Department of Psychiatry, Alpert Medical School.

Dr Zimmerman jest kierownikiem/dyrektorem Psych Products Press. Dr Young, dr Chelminski i dr Dalrymple nie zgłaszają żadnych mogących powodować konflikt interesów zależności finansowych ani powiązań z organizacjami komercyjnymi.

Adres do korespondencji: Mark Zimmerman, MD, Baryside Medical Center, 235 Plain St, Providence, RI 02905, Stany Zjednoczone; e-mail: mzimmerman@lifespan.org

na dowodach wiąże się z lepszymi wynikami? Czy niektórzy lekarze są skuteczniejsi niż inni oraz, zakładając, że istnieją różnice w wynikach pacjentów leczonych przez różnych lekarzy, czy jest możliwe poprawienie wyników pacjentów leczonych przez lekarzy osiągających wyniki leczenia poniżej przeciętnej? Na każde z tych i inne pytania można odpowiedzieć tylko pod warunkiem mierzenia wyników leczenia i włączenia tych pomiarów do rutynowej praktyki klinicznej.

Przez ostatnie 15 lat autorzy klinicznej grupy badawczej prowadzili projekt Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS).^{2,3} Zgodnie z sugestią zawartą w nazwie najważniejszym celem projektu MIDAS była poprawa jakości leczenia. Aby osiągnąć ten cel, przyjęli podstawowe założenie, zgodnie z którym niemożliwa jest poprawa jakości bez mierzenia wyników. Dlatego opracowali skale możliwe do wykorzystania w praktyce klinicznej.^{4,6} Pierwszym etapem projektu było opracowanie metod mierzenia wyników, które mogłyby zostać łatwo wykorzystane przez klinicystów pracujących poza tym programem. W tym artykule opisano praktyczne problemy, z którymi konfrontowano się i próbowano je rozwiązywać przy opracowywaniu programu oceny behawioralnych wyników leczenia.

Potencjalne przeszkody w mierzeniu wyników

W tabeli 1 wymieniono sześć przeszkód we wprowadzaniu pomiarów wyników leczenia w praktyce klinicznej, które trzeba pokonać.

Akceptacja pacjenta. Jeżeli pomiary są zbyt uciążliwe dla pacjentów, mogą oni uznać leczenie za niesatysfakcjonujące i zrezygnować z niego lub poszukać pomocy gdzie indziej.

Akceptacja lekarza. Jeżeli pomiary zaburzają codzienny rytm pracy lekarza, jest mniej prawdopodobne, że będzie je stosować.

Kliniczna przydatność pomiarów. Pomiary, które są istotne klinicznie i wpływają na poprawę efektywności przeprowadzania ocen klinicznych, częściej zostaną zaadapto-

wane niż narzędzia, które nie są przydatne przy podejmowaniu decyzji klinicznych.

Koszty. Narzędzia związane z większymi kosztami nabycia lub kosztami wykorzystującego je personelu pomocniczego są wykorzystywane rzadziej.

Niejednorodność diagnostyczna. Większość klinicystów leczy chorych z różnymi rozpoznaniem. Pomiary swoiste diagnostycznie będą mieć zastosowanie tylko u ograniczonej liczby pacjentów, dlatego należy wykorzystywać skale, które odnoszą się do osób z różnymi rozpoznaniem.

Gromadzenie danych. Jeżeli celem mierzenia wyników jest poprawa jakości leczenia, potrzebne będzie gromadzenie informacji dotyczących wyników uzyskiwanych w różnych grupach pacjentów w celu określenia wpływu zmiany sposobu opieki.

Każdą z tych przeszkód omówiono poniżej.

Problem heterogenności i koncepcją psychiatrycznych objawów życiowych

Opisując podejście terapeutyczne oparte na pomiarach przy nawiązaniu do leczenia stosowanego w badaniu STAR*D (Sequenced Treatment Alternatives to Relieve Depression), Trivedi i wsp.⁷ skupili swoją uwagę na chorych na depresję. Większość psychiatrów leczy niejednorodne grupy pacjentów z różnymi rozpoznaniem. DSM-IV⁸ wymienia ponad 400 rozpoznań zebranych w 17 kategoriach diagnostycznych. Niejednorodność diagnostyczna stanowi wyzwanie przy adaptowaniu terapii opartej na pomiarach w praktyce klinicznej.

Są trzy możliwe podejścia do rozwiązywania problemu heterogeniczności w mierzeniu wyników w praktyce klinicznej. Jedno z nich zakłada wykorzystywanie ogólnych narzędzi, które byłyby odpowiednie dla wszystkich pacjentów, niezależnie od rozpoznania. Narzędzie może odzwierciedlać zmiany w stanie klinicznym, jak np. Clinical Global Index (CGI) of Improvement,⁹ służące do całościowej oceny poprawy lub pogorszenia stanu pacjenta podczas leczenia. Alternatywnie, można wykorzystywać narzędzie zapewniające

nieswoistą pod względem diagnostycznym ocenę objawów lub funkcjonowania, np. Global Assessment of Functioning (GAF).⁸ Ograniczeniem tego typu podejścia jest niedostarczanie przez takie skale informacji dotyczących konkretnych objawów, na podstawie których często podejmowane są decyzje terapeutyczne. Drugą, przeciwną strategią będzie wykorzystywanie swoistych diagnostycznie skal oceny objawów. Dla większości częstych zaburzeń psychicznych opracowano i wystandaryzowano takie skale. Zaktualizowany ostatnio *Handbook of Psychiatric Measures*¹⁰ stanowi kompendium opisujące wiele z tych narzędzi. Chociaż idealne podejście do monitorowania wyników powinno obejmować stosowanie pomiarów swoistych w stosunku do określonego zaburzenia, wykorzystywanie wielu skal jest czasochłonne, a autorzy tego artykułu są sceptyczni co do tego, czy wielu lekarzy będzie z nich korzystać w rutynowej praktyce klinicznej. Trzecia strategia, którą autorzy tutaj proponują, polega na rutynowej ocenie dwóch osiowych, często występujących objawów psychicznych, analogicznie do rutynowego badania somatycznych objawów życiowych. Dlatego nazwano je „psychiatrycznymi objawami życiowymi” (psychiatric vital signs). Autorzy proponują uznanie za psychiatryczne objawy życiowe objawów depresji i lęku.

Somatyczne objawy życiowe odzwierciedlają podstawowe funkcje fizjologiczne, które są rutynowo oceniane w ośrodkach medycznych. Określenie „objawy życiowe” wiąże się z tym, że objawy te oceniają funkcje będące podstawowymi wskaźnikami życia. Nie sugerowano, żeby ocena objawów depresji i lęku była analogiczna w tym sensie. Chodzi raczej analogię z rutynowym pomiarem temperatury ciała, ciśnienia tętniczego, częstości akcji serca i oddechowemu na początku wizyty ambulatoryjnej, niezależnie od głównej skargi pacjenta. Na tej podstawie wykorzystywane jest określenie „objawy życiowe”.

Somatyczne objawy życiowe są czasami miarą pierwszorzędową (np. ciśnienie tętnicze u chorych leczonych z powodu nadciśnienia), a w innych przypadkach są miarą uzupełniającą. Podobnie, ocena nasilenia depresji i lęku może być pomiarem pierwszorzędowym lub uzupełniającym. Ponieważ zaburzenia nastroju i zaburzenia lękowe należą do najczęstszych zaburzeń psychicznych, rutynowa ocena objawów depresji i lęku może być uznana za podstawowy element badania, który dostarcza istotnych klinicznie, ilościowych informacji dotyczących skuteczności leczenia większości pacjentów z zaburzeniami psychicznymi. Oczywiście ocena objawów lęku i depresji nie wyklucza wykorzystywania dodatkowych, swoistych wobec konkretnych zaburzeń, narzędzi oceny. Biorąc jednak pod uwagę, jak rzadko psychiatrzy

TABELA 1
Potencjalne przeszkody we wprowadzaniu programu oceny wyników leczenia do praktyki klinicznej

Akceptacja pacjenta
Akceptacja klinicysty
Kliniczna przydatność pomiarów
Koszty
Niejednorodność diagnostyczna
Gromadzenie danych

Zimmerman M, Young D, Chelminsky I, Darymple K, *Primary Psychiatry*. Vol 17, No 7. 2010.

wykorzystują obecnie wystandaryzowane skale do oceny wyników, autorzy są przekonani, że zaadaptowanie tych dwóch wskaźników wyników leczenia posunie naprzód wysiłki zmierzające do wprowadzenia opieki psychiatrycznej opartej na mierzeniu.

W ramach projektu MIDAS autorzy szacowali częstość objawów depresji i lęku w grupie 3000 leczonych ambulatoryjnie pacjentów psychiatrycznych, ocenianych za pomocą Ustrukturyzowanego wywiadu klinicznego dla DSM-IV; SCID (Structured Clinical Interview for DSM-IV).¹¹ SCID uzupełniano pozycjami do oceny objawów pochodzącyymi z Schedule for Affective Disorders and Schizophrenia (SADS),¹² z których dwie służyły do oceny obniżonego nastroju i lęku. Na podstawie oceny w SADS stwierdzono, że 79,3% ($n=2378$) pacjentów zgłaszało istotne klinicznie objawy depresji o co najmniej łagodnym nasileniu, 64,4% ($n=1932$) zgłaszało lęk o co najmniej łagodnym nasileniu, zaś 87,4% ($n=2621$) zgłaszało lęk lub depresję. Dane w tabeli 2 ilustrują dużą częstość występowania lęku i depresji u chorych zaliczanych do wszystkich kategorii diagnostycznych. Uzyskane wyniki przemawiają zatem za rutynową oceną objawów lęku i depresji w ambulatoryjnej praktyce klinicznej, ponieważ prawie wszyscy pacjenci będą zgłaszać te problemy. Nawet w przypadku chorych bez depresji lub lęku trzeba stwierdzić, że ocena tych objawów jest istotna, podobnie jak ocena niektórych parametrów fizjologicznych w praktyce medycznej, takich jak ciśnienie tętnicze czy temperatura ciała, niezależnie od powodu wizyty.

Podejście do oceny objawów oparte na koncepcji „objawów życiowych” odnosi się nie tylko do problemu, co mierzyć, ale również, kiedy mierzyć. W ośrodkach ambulatoryjnych można rozważyć dwa podejścia dotyczące częstości i czasu dokonywania pomiarów – stałe punkty czasowe *versus* ocena podczas każdej wizyty. W ramach tego pierw-

szego podejścia oceny dokonuje się w określonych wcześniej punktach czasowych, np. na początku, a potem co 3 miesiące, na początku i na końcu leczenia (tzn. przed leczeniem i po jego zakończeniu). Praktyczny problem związany z tym podejściem wiąże się z tym, że pacjenci często bez wcześniejszego uprzedzenia rezygnują z leczenia, co prowadzi do utraty istotnych informacji na temat ich stanu klinicznego w momencie zakończenia terapii. Innym praktycznym problemem jest potrzeba śledzenia, kiedy przeprowadzić kolejną ocenę. Co więcej, podejście to nie wprost sugeruje, że określenie stanu pacjenta nie jest podstawowym elementem wszystkich wizyt terapeutycznych. Zalecenie, aby mierzenia wyników dokonywać tylko w określonych punktach czasowych, sugeruje, że takie oceny są uciążliwe, zatem w celu ograniczenia związanego z nimi obciążenia nie należy dokonywać ich rutynowo. Postrzeganie pomiarów wyników jako czegoś uciążliwego będzie zaburzać rozpowszechnienie się tej procedury. Z drugiej strony, zgodnie z podejściem „podczas każdej wizyty” czy „psychiatrycznych objawów życiowych”, oceny dokonuje się w czasie wszystkich wizyt kontrolnych i uważa się to za integralny element leczenia. Aby się upewnić, lekarze już teraz rutynowo oceniają wyniki podczas wizyt kontrolnych, chociaż czynią to w niewystandaryzowany, nieilościowy sposób. Dzięki dostępności krótkich, chociaż przy tym rzetelnych i trafnych skal do oceny objawów depresji i lęku realne jest włączenie oceny powyższych parametrów do pomiaru „medycznych objawów życiowych”.

Akceptowalność pomiaru wyników

Nie ma ograniczeń w wykorzystywaniu parametrów do monitorowania wyników leczenia. Podczas podejmowania decyzji o wyborze narzędzia największe znaczenie mają

dwie perspektywy – pacjenta i lekarza. Chory powinni uznać dane narzędzie za przyjazne dla użytkownika, a instrukcje za łatwe do wypełniania. Pytania powinny być zrozumiałe i istotne dla problemu pacjenta. Skale powinny być krótkie, możliwe do wypełnienia w czasie nieprzekraczającym 2-3 minut, tak, aby chory podczas rutynowego zapisywania się na wizyty kontrolne nie był narażony na niewygodę związaną z koniecznością przychodzenia 10-15 minut wcześniej w celu wypełnienia skali.

Stosowane narzędzie powinno dostarczać lekarzowi przydatnych z klinicznego punktu widzenia informacji oraz poprawiać skuteczność przeprowadzania oceny klinicznej, a zatem mieć praktyczną wartość dla lekarza. Oczywiście klinicyści muszą być w stanie zaufać informacjom dostarczanym przez skalę, którą wykorzystują. Dlatego parametr oceny wyników powinien opierać się na solidnych podstawach psychometrycznych, charakteryzować się dobrą rzetelnością, trafnością i czułością na zmianę. Lekarze i poradnie powinni również uznać, że wykorzystywane narzędzie jest przyjazne dla użytkownika; powinno być łatwe w wykorzystaniu i obliczaniu wyników po minimalnym przeszkoleniu.

Grupa autorów¹³ oceniała łatwość i akceptowalność wykorzystywania kwestionariusza samooceny depresji w celu oceny wyników leczenia w rutynowej praktyce klinicznej w dwóch badaniach dotyczących cierpiących na depresję psychiatrycznych pacjentów ambulatoryjnych w trakcie leczenia. Pacjenci wypełniali kwestionariusz oceniający, jak obciążające było dla nich wypełnienie skali podczas wizyty (0 – bardzo mało obciążające, 3 – bardzo obciążające) oraz gotowość do wypełniania skali podczas każdej wizyty, aby pomóc w monitorowaniu postępów w leczeniu (0 – wcale nie chce, 3 – bardzo chce wypełniać podczas każdej wizyty). Prawie wszyscy pacjenci wypełnianie kwestionariusza uznali za bardzo mało lub tylko trochę

TABELA 2
Częstość istotnych klinicznie objawów depresji i lęku u 3000 leczonych ambulatoryjnie pacjentów psychiatrycznych z podstawowymi rozpoznaniem z różnych kategorii diagnostycznych

Kategoria diagnostyczna	<i>n</i>	Depresja % (<i>n</i>)	Lęk % (<i>n</i>)	Depresja i lęk % (<i>n</i>)
Zaburzenia nastroju	1702	90,9 (1547)	66,4 (1130)	93,9 (1598)
Zaburzenia lękowe	546	71,2 (389)	77,5 (423)	89,2 (487)
Zaburzenia spowodowane używaniem substancji	56	53,6 (30)	46,4 (26)	71,4 (40)
Zaburzenia psychotyczne	51	54,9 (28)	62,7 (32)	74,5 (38)
Zaburzenia pod postacią somatyczną	65	70,8 (46)	66,2 (43)	81,5 (53)
Zaburzenia kontroli impulsów	53	64,2 (34)	50,9 (27)	73,6 (39)
Zaburzenia odżywiania się	20	75,0 (15)	55,0 (11)	85,0 (17)
Zaburzenia adaptacyjne	180	70,0 (126)	53,9 (97)	83,3 (150)
Zaburzenia uwagi	77	41,6 (32)	40,3 (31)	54,5 (42)
Zaburzenia osobowości	92	72,8 (67)	53,3 (49)	79,3 (73)

Zimmerman M, Young D, Chelminsky I, Darymple K. *Primary Psychiatry*. Vol 17, No 7. 2010.

obciążające (98%, $n=49$), zaś żaden pacjent nie uznał tej czynności za bardzo obciążającą. Ponad 90% pacjentów zaznaczyło gotowość do wypełniania skali podczas każdej wizyty w przyszłości, jeżeli ich lekarz uzna to za pomocne (94%, $n=47$).

Oczywiście, akceptowalność i łatwość wykorzystania mogą różnić się w zależności od skali. Niektóre narzędzia składają się z ≥ 100 stwierdzeń, natomiast inne zawierają ich mniej niż 10. W podsumowanym powyżej badaniu wykorzystywano skalę opracowaną przez autorów w ramach projektu MIDAS – Clinically Useful Depression Outcome Scale (CUDOS),^{5,14} krótkie narzędzie, w którym respondentów prosi się o ocenę 16 objawów w 5-punktowej skali Likerta. W drugim badaniu oceniającym łatwość wykorzystania,¹⁵ oddzielna grupa 50 leczonych ambulatoryjnie chorych na depresję wypełniała jednocześnie skalę CUDOS i skalę depresji Becka (Beck Depression Inventory, BDI) podczas wizyty kontrolnej. W przeciwieństwie do skali CUDOS, w BDI respondenci muszą przeczytać grupę czterech stwierdzeń i wybrać stwierdzenie, które ich zdaniem najlepiej określa ich samopoczucie w ciągu ostatniego tygodnia. Zatem BDI zawiera więcej informacji do przeczytania w porównaniu z CUDOS. Po wypełnieniu tych dwóch kwestionariuszy pacjenci wypełniali jeszcze jeden, w którym opisywali które z tych dwóch narzędzi zajmował im mniej czasu, które z nich było bardziej zrozumiałe, mniej obciążające podczas wypełniania i które byłoby łatwiejsze do zaakceptowania w razie konieczności wypełniania podczas każdej wizyty kontrolnej. Istotnie więcej pacjentów stwierdziło, że wypełnienie skali CUDOS zajmowało mniej czasu niż wypełnienie BDI (64 vs 12%, $p < 0,05$) i było mniej obciążające (50 vs 10%, $p < 0,05$). Prawie trzykrotnie więcej badanych zaznaczyło, że woleliby wypełniać skalę CUDOS niż skalę BDI w celu monitorowania wyników leczenia (40 vs 14%, $p < 0,05$). Wyniki te sugerują, że badani nie uznali wypełniania skali za obciążające, szczególnie jeżeli była krótka. Oprócz skali CUDOS, innym krótkim narzędziem do oceny depresji, które można z łatwością wykorzystywać w rutynowej praktyce klinicznej jest złożony z 9 pozycji Patient Health Questionnaire.¹⁶

Problem gromadzenia danych

Na początku tego artykułu postawiono pytanie, czy chorzy leczeni w ośrodkach medycznych wysoko ocenionych w corocznym rankingu *U.S. News and World Report* częściej osiągają dobre wyniki leczenia. Aby odpowiedzieć na to pytanie trzeba nie tylko mierzyć wyniki, ale również gromadzić dane pochodzące od różnych chorych i lekarzy. Powyżej

autorzy zwrócili uwagę na problem gromadzenia danych. Chociaż gromadzenie danych nie jest niezbędne do uświadamiania sobie korzyści klinicznych w leczeniu opartym na pomiarach u indywidualnych pacjentów, to gromadzenie danych jest niezbędne w dążeniu do poprawy jakości leczenia, kiedy porównywane są wyniki leczenia przed zmianą sposobu opieki i po niej. Proces gromadzenia danych może być intensywną, a zatem kosztowną, pracą. Aby pokonać potencjalne przeszkody związane z kosztami wpisywania danych, autorzy opracowali sieciowy system zarządzania danymi i obliczania wyników w skali CUDOS oraz gromadzenia danych pochodzących od różnych klinicystów w celu określenia częstości reakcji na leczenie i remisji po zakończeniu ostrej fazy leczenia.

Ocena sieciowa

Dostępny w sieci program zarządzania oceną wyników ma pod kilkoma względami przewagę nad techniką oceny za pomocą kartki i ołówka (tab. 3). Dostępne w sieci skale mogą być wypełniane przez chorych w domu, co nie wymaga od nich wcześniejszego przyjeżdżania na wizytę lub pozostawiania po jej zakończeniu w celu wypełnienia skali. Sondaż zarządzany przez komputer może zachęcać respondentów, aby upewnili się, czy odpowiedzieli na wszystkie pytania, zmniejszając dzięki temu liczbę utraconych danych. System internetowy umożliwia zmniejszenie kosztów administracyjnych związanych z kopiowaniem i rozdawaniem papierowych kwestionariuszy oraz liczeniem wyników. Wysokie koszty związane z tworzeniem i utrzymywaniem baz danych w celu oceny wyników w dużej grupie pacjentów na podstawie zarządzania, obliczania wyników i wprowadzania danych z kwestionariuszy papierowych również ulegają zmniejszeniu dzięki wykorzystywaniu systemu sieciowego. Co więcej, ponieważ dane zbierane za pośrednictwem internetu, są automatycznie przekazywane do bazy danych, zmniejszają się błędy związane z wprowadzaniem danych.¹⁷

Nie można założyć, że wykorzystywanie papierowych i komputerowych wersji tego samego testu pozwoli na uzyskanie równoważnych wyników.¹⁸⁻²⁰ Mimo analogicznej za-

wartości różnice dotyczące formatu mogą wpływać na wyniki.^{21,22} Na przykład w niektórych testach komputerowych poszczególne pozycje skali pojawiają się pojedynczo na ekranie, natomiast testy papierowe prezentują na jednej stronie wiele pozycji. Prezentowanie pozycji pojedynczo, a nie grupowo może prowadzić do różnic w zwracaniu uwagi na poszczególne pozycje, a wtórnie – do odmiennych korelacji pozycja-skala. Niektóre komputerowe wersje testów nie umożliwiają powrotu do poprzednich pozycji, natomiast testy papierowe umożliwiają zmianę udzielonych wcześniej odpowiedzi. Może to prowadzić do frustracji respondenta przy korzystaniu z wersji komputerowej, wpływając przez to na odpowiedzi na pozostałe pytania w skali. Respondenci mogą również ponownie rozważać wcześniejsze odpowiedzi po udzieleniu odpowiedzi na kolejne pytania. Niektóre skale komputerowe nie pozwalają na opuszczanie pytań. „Hipoteza szczerości” zakłada, że respondenci są bardziej prawdomówni, kiedy odpowiadają na testy w komputerze. Każdy z tych czynników mógłby wpływać na psychometryczne właściwości testu, a także na wartości uznawane za „punkty odcięcia”, wykorzystywane do określania „przypadków” (lub osób w remisji). Dlatego konieczne jest wykazanie równoważności.^{18,19}

Wykorzystywanie skal dostępnych przez internet wiąże się z dodatkowym potencjalnym źródłem błędów i zmienności, ponieważ pacjenci mogą wypełniać kwestionariusze podczas oglądania telewizji, rozmawiania przez telefon, jedzenia, czy angażowania się w innego rodzaju rozpraszające aktywności.²³ Niewystandaryzowane warunki, w jakich mogą być wypełniane skale dostępne przez internet mogą zmieniać udzielane odpowiedzi, co wymaga potwierdzenia ich równoważności w wyniku porównaniu z badaniem przeprowadzonym w gabinecie.¹⁷

Mimo potencjalnych źródeł błędów, badania porównujące wykorzystywanie skal w postaciach papierowych, komputerowych, w tym dostępnych przez internet, wykazały duży stopień zgodności między tymi dwiema metodami oceny.²⁴⁻²⁷ W niewielu badaniach oceniano wykorzystywanie skali dostępnych w sieci. Autorzy nie znają żadnego

TABELA 3
Potencjalne zalety oceny przez internet

Wygoda pacjenta
Ograniczenie utraconych danych
Zmniejszenie kosztów
Automatyczne obliczanie wyników
Automatyczne tworzenie baz danych
Gromadzenie danych pochodzących od różnych pacjentów

Zimmerman M, Young D, Chelminski I, Dalrymple K. *Primary Psychiatry*. Vol 17, No 7. 2010.

badania dotyczące pacjentów psychiatrycznych leczonych w rzeczywistych warunkach. Należy również zachować ostrożność przed ekstrapolowaniem wyników skal dostępnych w formie papierowej na wyniki skal dostępnych w sieci. Dlatego należy kontynuować wysiłki zmierzające do wykazania ich równoważności. W ramach projektu MIDAS autorzy przeprowadzili pierwsze badania porównujące papierową i internetową postać skali oceny depresji u chorych na depresję w ciągłym leczeniu ambulatoryjnym.

Autorzy prezentują tutaj wstępne wyniki prowadzonego obecnie badania porównującego papierową i sieciową wersję skali CUDOS. Na czterdzieści osiem godzin przed wizytą 30 pacjentów z depresją objętych ciągłą opieką ambulatoryjną otrzymało e-mail przypominający o spotkaniu, w którym poproszono ich także o wypełnienie dostępnej w sieci wersji skali CUDOS (CUDOS-W). Pod koniec wizyty pacjentów proszono o wypełnienie papierowej wersji skali CUDOS. Wyjaśniano im, że okoliczności i sposób wypełniania skali wpływają czasami na udzielane odpowiedzi, dlatego ważne jest zbadanie stopnia porównywalności komputerowej i papierowej wersji skali. Podczas wizyty klinicysta wypełniał Montgomery-Asberg Depression Rating Scale (MADRS)²⁸ oraz dokonywał oceny w skalach GAF i CGI-nasilenie (CGI-S, Severity).²⁹

Formaty papierowej i internetowej wersji skali CUDOS były identyczne. Wykorzystywanie skali dostępnej w sieci polega raczej na jednoczesnym prezentowaniu całej skali niż wyświetlaniu poszczególnych pozycji kolejno po sobie. Pacjenci mogą zmieniać udzielone wcześniej odpowiedzi. Kwestionariusz nie może zostać wysłany zanim nie zostaną wypełnione wszystkie pozycje.

Badana próba obejmowała 6 mężczyzn (20%) i 24 kobiety (80%) w wieku 22-85 lat (M 47,0, odchylenie standardowe [SD] 12,2). Średni wynik w skali MADRS (11,9, SD 12,3) i CGI-S (1,2, SD 1,3) wskazywał na łagodne nasilenie objawów depresyjnych. Średni wynik w skali GAF wynosił 66,1 (SD 10,2).

Przeciętny odstęp czasowy między wypełnieniem wersji papierowej i dostępnej przez internet wynosił 1,1 dnia (SD 0,8). Korelacja między CUDOS a CUDOS-W była wysoka (ICC 0,95, $p < 0,001$). Średnie wyniki uzyskane w wersji papierowej i internetowej były podobne ($18,9 \pm 14,5$ vs $18,5 \pm 13,8$, sparowane $t = 0,4$, NS). We wcześniejszych badaniach oceniających rzetelność papierowej wersji skali CUDOS autorzy stwierdzili, że punkt odciążenia dla pacjentów będących w remisji wynosił 20 punktów. Na podstawie tego punktu odciążenia stwierdzono zgodność rzędu 94,7% między papierową i internetową wersją obu skal w określaniu remisji u pacjen-

tów ($k = 0,90$).

Wewnętrzna zgodność papierowej i internetowej wersji CUDOS była wysoka (współczynnik Cronbacha $\alpha = 0,94$ i $0,93$, odpowiednio). Dla każdej pozycji korelacja między CUDOS i CUDOS-W była istotna (mediana 0,86).

Obydwie, papierowa i internetowa, wersje skali CUDOS były istotnie skorelowane w wynikami skali MADRS ($r = 0,92$ i $0,89$, odpowiednio); CGI-S ($r = 0,94$ i $0,93$, odpowiednio) oraz GAF ($r = 0,94$ i $0,90$, odpowiednio). Żadna z różnic dotyczących korelacji między wersjami papierową i internetową oraz trafności skal nie była istotna statystycznie.

Wstępne wyniki trwającego obecnie badania przemawiają za rzetelnością i trafnością internetowej wersji skali CUDOS. Wewnętrzna zgodność, korelacje pozycja-skala i korelacje w zewnętrznymi skalami odniesienia były podobnie wysokie dla wersji internetowej i papierowej. Strona internetowa, na której można znaleźć skalę CUDOS i opracowaną przez autorów Clinically useful Anxiety Outcome Scale,⁴⁶ jest dostępna pod adresem www.outcometracker.org.³⁰ Lekarze mogą teraz korzystać z niej za darmo.

W jaki sposób ocean wyników może wpłynąć na poprawę opieki nad chorymi na depresję?

Autorzy są przekonani, że obecnie dostępne są narzędzia umożliwiające lekarzom włączenie do pracy ambulatoryjnej metod oceny wyników w sposób niezaburzający codziennej pracy. Dlaczego lekarze powinni zmieniać swoje zachowanie i wykorzystywać w leczeniu podejście oparte na pomiarach? Autorzy chcieliby móc wskazać badania empiryczne wykazujące korzyści terapeutyczne z leczenia opartego na ocenie wyników. W rzeczywistości dostępnych jest niewiele danych potwierdzających, że taki rodzaj leczenia wpływa na poprawę wyników. Badacze uczestniczący w badaniu STAR*D utrzymywali, że leczenie oparte na pomiarach było korzystne, ponieważ częstość reakcji na leczenie w I fazie badania STAR*D była podobna do obserwowanej przeważnie w sponsorowanych przez firmy farmaceutyczne badaniach nad skutecznością leków, pomimo o wiele szerszych kryteriów włączenia do udziału w badaniu STAR*D, wskutek czego uczestniczyło w nim więcej pacjentów „trudnych do leczenia”, z zaburzeniami współistniejącymi.⁷

Z kilku powodów, z których leczenie oparte na pomiarach może wpływać na poprawę wyników leczenia. Jednym z nich jest lepsze wykrywanie objawów rezydualnych u pacjentów, których stan poprawił się podczas leczenia, jednak nie udało się uzyskać re-

misji. W wielu badaniach wykazano, że obecność objawów rezydualnych u chorych na depresję, którzy uzyskali poprawę w trakcie leczenia, prognozuje gorsze odległe wyniki.³¹⁻³⁴ W jakim stopniu lekarze wykrywają te objawy rezydualne? Autorzy nie znają badań, w których poszukiwano by odpowiedzi na to pytanie. Jak jednak wykazano w badaniu STAR*D, objawy rezydualne występują u 33% pacjentów doświadczających łagodnych do umiarkowanych objawów pod koniec ostrej fazy leczenia citalopramem.⁷ Chociaż prawdą jest, że częstość remisji w ostrej fazie leczenia w badaniu STAR*D była niewielka, pomimo wykorzystywania oceny wyników leczenia przy podejmowaniu decyzji terapeutycznych, łączny odsetek remisji po wielu poziomach leczenia wynosił >60%.^{35,36} Autorzy zgadzają się z przypuszczeniem badaczy uczestniczących w STAR*D, że ilościowa ocena wpływała korzystnie na wyniki leczenia, ponieważ nie dało się ignorować niepełnej odpowiedzi. Wykorzystywanie narzędzi do pomiarów powoduje, że objawy rezydualne, które narażają pacjentów na większe ryzyko nawrotu, rzadziej pozostają nierozpoznane, a także zmniejsza prawdopodobieństwo błędnej konkluzji, że pacjenci, u których utrzymują się objawy pozytywnie zareagowali na leczenie, przyjmowanej za prawdziwą na podstawie ogólnych stwierdzeń, że czują się „w porządku” lub „okay”.

Ważne jest również mierzenie wyników od początku leczenia. W kilku badaniach wykazano, że pacjenci, którzy nie uzyskali poprawy o co najmniej 20% w ciągu 2 tygodni rzadziej uzyskują poprawę po 6-10 tygodniach.^{37,38} Brak reakcji na leczenie na wczesnym etapie jest najlepszym predykatorem braku reakcji na leczenie lub remisji i dlatego może sugerować lekarzowi potrzebę zmiany leku. Należy jednak zachować ostrożność, ponieważ w żadnym badaniu nie wykazano, aby zmiana leczenia po dwóch tygodniach u osób niereagujących prowadziła do poprawy wyników terapii.

Wielu chorych na depresję wypada z leczenia w ciągu jego pierwszych 6 miesięcy,^{39,40} a wielu jest leczonych przez wiele lat ze względu na przewlekły charakter choroby. Rutynowa ocena wyników leczenia za pomocą skali samooceny może z różnych powodów poprawić skuteczność terapeutyczną, w zależności od stadium leczenia. Wypełnianie skal samooceny zwiększa aktywne uczestnictwo pacjentów w terapii, co może ułatwiać ich włączanie się w inne aktywności terapeutyczne, takie jak ćwiczenia fizyczne lub przyjemne aktywności. Pacjenci, którzy są aktywniejsi w procesie leczenia, i tacy, którzy są przekonani, że ich lekarze lepiej rozumieją ich stan kliniczny, mogą częściej utrzymywać się w leczeniu. Trafna ocena objawów może pomóc lekarzom w identyfikowaniu tych ob-

szarów poprawy, które wcześniej nie zostały wykryte. Wyobraźmy sobie na przykład chorego, który ciągle jest w obniżonym nastroju, pesymistycznie nastawiony, niezmotywowany i deprecjonujący siebie, a na początku wizyty kontrolnej stwierdza, że nie czuje się lepiej, ale w rzeczywistości śpi lepiej, czuje nieco więcej energii, lepiej się odżywia i koncentruje. Wykrywanie pewnych obszarów poprawy może zmniejszyć nihilizm terapeutyczny pacjentów, sprzyjając w ten sposób kontynuowaniu leczenia. Dlatego dokładniejsza ocena objawów mogłaby nie tylko poprawić wykrywanie objawów rezydualnych u pacjentów, którzy relacjonują, że czują się lepiej, ale również poprawić wykrywanie niewielkiej poprawy (rzędu 20-30% w porównaniu z poziomem wyjściowym), co może być zwiastunem dalszej poprawy u pacjentów, którzy jeszcze nie czują się dobrze.^{37,38}

Chorzy obserwowani prospektywnie przez wiele lat mogą wyjątkowo skorzystać z rutynowego wykorzystywania skal. Patrząc na wykresy sporządzane na podstawie uzyskanych wyników można, na przykład, łatwiej wykryć sezonowe wahania nasilenia objawów. Chory, u którego dochodzi do nawrotu i który z powodu zależnych od stanu depresji zniekształcając poznawczych błędnie ocenia skuteczność leczenia wskutek minimalizowania lub pominięcia okresów utrzymującej się remisji, może być bardziej otwarty na weryfikowalne doniesienia dotyczące przebiegu jego choroby, kiedy konfrontuje się go z wypełnianymi przez niego w przeszłości formularzami świadczącymi o minimalnym nasileniu objawów.

Istnieje zatem wiele uzasadnień teoretycznych tego, że leczenie oparte na pomiarach może wpłynąć na poprawę wyników. Chociaż nie przeprowadzono badań, które wykazałyby korzyści z leczenia opartego na ocenie wyników, koszty zaadaptowania takiego podejścia są tak niewielkie, a potencjalne korzyści tak istotne, że trudno jest usprawiedliwić niepodjęcie próby wprowadzenia takiego podejścia na pewien czas.

Jak sprawić, aby klinicyści zaadaptowali leczenie oparte na ocenie wyników

Ten artykuł był adresowany do praktykującego lekarza. Próbowano znaleźć w nim racjonalne uzasadnienie dla modyfikowania opieki klinicznej w kierunku uwzględniania ocen ilościowych. Niektórzy klinicyści będą przeprowadzać eksperymenty w celu określenia, czy będą zadowoleni z poprawy skuteczności lub wyników leczenia oraz kosztów. Czy pacjenci będą zniechęceni wypełnianiem skal? Czy będą protestować, wyrażać niezadowolenie, czy też zrobią to chętnie? Czy wpłynie to na zakłó-

cenie rytmu pracy?

Również czynniki ekonomiczne mogą wpływać na ułatwienie zmiany sposobu postępowania w kierunku uwzględniania leczenia opartego na pomiarach. Polityka finansowa towarzystw ubezpieczeniowych i pracodawców może zachęcać klinicystów do zmiany. W jaki sposób towarzystwo ubezpieczeniowe może sprawić, żeby lekarze zaczęli wykorzystywać skalę własnego wyboru u wszystkich pacjentów w celu oceny wyników leczenia? Nie mogą tego żądać, czy wymagać, ponieważ taka postawa prawdopodobnie wywołałaby opór.

Zamiast narzuconej z góry zmiany postępowania, towarzystwa ubezpieczeniowe mogą dostarczać bodźców zachęcających do oceny wyników leczenia. Takimi bodźcami mogą być: kierowanie odsyłaniem pacjentów, ograniczanie poziomu „mikroleczenia” i nadzoru, zwiększenie refundacji kosztów lub objęcie refundacją elektronicznych medycznych systemów zapisu danych. Autorzy zakładaliby, że podejście oparte na motywowaniu i współpracy bardziej sprzyja odniesieniu sukcesu.

Podobnie pracodawcy, czyli nabywcy ubezpieczeń zdrowotnych, mogą zachęcać towarzystwa oferujące ubezpieczenia medyczne do kontrolowania, w jaki sposób wydawane są ich pieniądze. Jest to możliwe tylko pod warunkiem oceny wyników i gromadzenia danych pochodzących od dużych grup pacjentów.

Innym czynnikiem motywującym jest organizacja systemu opieki zdrowotnej. Pierwsza organizacja w danym rejonie, która zaadaptuje leczenie oparte na pomiarach, może być w stanie współzawodniczyć o lepszy kontrakt. Takie organizacje mogą uzyskane przez siebie dane dotyczące wyników leczenia i jakość opieki przedstawiać opinii publicznej, a także ubezpieczycielom. Duże praktyki medyczne mogą spowodować, że lekarze zastosują leczenie oparte na pomiarach, jeżeli będą mieć wrażenie, że ich rentowność finansowa w przyszłości jest uzależniona od opracowania programu oceny wyników leczenia.

Podsumowanie

Autorzy przewidują, że w ciągu następujących kilku lat coraz większa liczba lekarzy i programów instytucjonalnych będzie rozważać wprowadzenie wystandaryzowanych narzędzi oceny do swojej praktyki. Ostatecznym celem programu oceny wyników jest poprawa jakości leczenia. Aby program oceny wyników leczenia był skuteczny, gromadzone informacje powinny być istotne klinicznie i możliwe do zliczenia, a metody zbierania danych nie powinny istotnie zaburzać rytmu pracy lekarza. Zbierane dane po-

winny być wykorzystywane do oceny wpływu różnych parametrów na wynik leczenia oraz tego, czy zmiany formy opieki wpływają na poprawę wyników leczenia. Obecnie istnieją narzędzia i techniki pozwalające na przewyższenie wyzwań związanych z zaadoptowaniem podejścia opartego na pomiarach, które, należy mieć nadzieję, przyspieszą włączenie mierzenia wyników do rutynowej praktyki klinicznej.

Piśmiennictwo

- America's Best Hospitals. U.S. *News and World Report*. 2009;140(7):89-112.
- Zimmerman M. Integrating the assessment methods of researchers in routine clinical practice: The Rhode Island Methods to Improve Diagnostic Assessment and Services (MIDAS) project. In: First MB, ed. *Annual Review of Psychiatry*. Washington, DC: American Psychiatric Press, Inc.; 2003:29-34.
- Zimmerman M, Mattia JL. Psychiatric diagnosis in clinical practice: is comorbidity being missed? *Compr Psychiatry*. 1999;40(3):182-191.
- Zimmerman M, Mattia JL. A self-report scale to help make psychiatric diagnoses: The Psychiatric Diagnostic Screening Questionnaire (PDSQ). *Arch Gen Psychiatry*. 2001;58(8):787-794.
- Zimmerman M, Chelminski I, McGlinchey JB, Posternak MA. A clinically useful depression outcome scale. *Compr Psychiatry*. 2008;49(2):131-140.
- Zimmerman M, Chelminski I, Young D, Dalrymple KL. A clinically useful anxiety outcome scale. *J Clin Psychiatry*. 2010;71(5):534-542.
- Trivedi M, Rush A, Wisniewski S, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28-40.
- Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
- Guy W. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health; 1976.
- Rush AJ, First MB, Blacker D. *Handbook of Psychiatric Measures*. 2nd ed. Washington, DC: American Psychiatric Publishing; 2008.
- First MB, Spitzer RL, Gibbon M, Williams JB. *Structured Clinical Interview for DSM-IV Axis I Disorders - Patient edition (SCID-I/P, version 2.0)*. New York, NY: Biometrics Research Department, New York State Psychiatric Institute; 1995.
- Endicott J, Spitzer RL. A diagnostic interview: the schedule for affective disorders and schizophrenia. *Arch Gen Psychiatry*. 1978;35(7):837-844.
- Zimmerman M, McGlinchey J. Depressed patients' acceptability of the use of self-administered scales to measure outcome in clinical practice. *Ann Clin Psychiatry*. 2008;20(3):125-129.
- Zimmerman M, Posternak M, Chelminski I. Using a self-report depression scale to identify remission in depressed outpatients. *Am J Psychiatry*. 2004;161(10):1911-1913.
- Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry*. 1961;4:561-571.
- Kroenke K, Spitzer R, Williams J. The PHQ-9. Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.
- Coles ME, Cook LM, Blake TR. Assessing obsessive compulsive symptoms and cognitions on the internet: evidence for the comparability of paper and internet administration. *Behav Res Ther*. 2007;45(9):2232-2240.
- Buchanan T. Online assessment: desirable or dangerous? *Prof Psychol Res Pr*. 2002;33:148-154.
- Buchanan T. Internet-based questionnaire assessment: appropriate use in clinical contexts. *Cogn Behav Ther*. 2003;32:100-109.
- McDonald A. The impact of individual differences on

- the equivalence of computer-based and paper-and-pencil educational assessments. *Comput Educ.* 2002; 39:299-312.
21. Webster J, Compeau D. Computer-assisted versus paper-and-pencil administration of questionnaires. *Behav Res Methods Instrum Comput.* 1996;28:567-576.
 22. *Guidelines for Computer-based Tests and Interpretations.* Washington, DC: American Psychological Association; 1986.
 23. Barbeite F, Weiss E. Computer self-efficacy and anxiety scales for an internet sample: Testing measurement equivalence of existing measures and development of new scales. *Comput Human Behav.* 2004;20:1-15.
 24. Ogles BM, France CR, Lunnen KM, Bell MT, Goldfarb M. Computerized depression screening and awareness. *Community Ment Health J.* 1998;34(1):27-38.
 25. Merten T, Ruch W. A comparison of computerized and conventional administration of the German versions of the Eysenck Personality Questionnaire and the Carroll Rating Scale for Depression. *Pers Individ Dif.* 1996;20:281-291.
 26. Peterson L, Johannsson V, Carlsson S. Computerized testing in a hospital setting: Psychometric and psychological effects. *Comput Human Behav.* 1996; 12:339-350.
 27. Cook IA, Balasubramani GK, Eng H, et al. Electronic source materials in clinical research: acceptability and validity of symptom self-rating in major depressive disorder. *J Psychiatr Res.* 2007;41(9):737-743.
 28. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry.* 1979;134:382-389.
 29. Guy W. *ECDEU Assessment Manual for Psychopharmacology.* Revised. Rockville, MD: National Institute of Mental Health; 1976.
 30. Outcome Tracker. Available at: www.outcometracker.org. Accessed May 28, 2010.
 31. Paykel ES, Ramana R, Cooper Z, Hayhurst H, Kerr J, Barocka A. Residual symptoms after partial remission: An important outcome in depression. *Psychol Med.* 1995;25(6):1171-1180.
 32. Thase ME, Simons AD, McGeary J, et al. Relapse after cognitive behavior therapy of depression: Potential implications for longer courses of treatment. *Am J Psychiatry.* 1992;149(8):1046-1052.
 33. Simons A, Murphy G, Levine J, Wetzel R. Cognitive therapy and pharmacotherapy for depression. Sustained improvement over one year. *Arch Gen Psychiatry.* 1986;43(1):43-48.
 34. Faravelli C, Ambonetti A, Pallanti S, Pazzagli A. Depressive relapses and incomplete recovery from index episode. *Am J Psychiatry.* 1986;143(7):888-891.
 35. Quitkin F, McGrath P, Stewart J, et al. Remission rates with 3 consecutive antidepressant trials: Effectiveness for depressed outpatients. *J Clin Psychiatry.* 2005;66(6): 670-676.
 36. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. *Am J Psychiatry.* 2006;163(11): 1905-1917.
 37. Szegedi A, Jansen WT, van Willigenburg AP, van der Meulen E, Stassen HH, Thase ME. Early improvement in the first 2 weeks as a predictor of treatment outcome in patients with major depressive disorder: a meta-analysis including 6562 patients. *J Clin Psychiatry.* 2009;70(3):344-353.
 38. Tadic A, Helmreich I, Mergl R, et al. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *J Affect Disord.* 2010;120(1-3): 86-93.
 39. Lin EH, Von Korff M, Katon W, et al. The role of the primary care physician in patients' adherence to antidepressant therapy. *Med Care.* 1995;33(1):67-74.
 40. Demyttenaere K, Enzlin P, Dewe W, et al. Compliance with antidepressants in a primary care setting. 1: beyond lack of efficacy and adverse events. *J Clin Psychiatry.* 2001;62 suppl 22:30-33.