

Tomasz Rzepiński

Ewidencja empiryczna w procesie podejmowania decyzji klinicznych

1. WSTĘP

Fakt, że decyzje kliniczne powinny posiadać uzasadnienie empiryczne nie budzi wątpliwości. Jakie jednak dane stanowią wiarygodną bazę empiryczną? W odniesieniu do terapii współczesne rozwiązanie tej kwestii zdominowane zostało przez powstałą w latach 90. XX wieku koncepcję *evidence-based medicine* (EBM).¹ Koncepcja ta określana jest mianem nowego paradygmatu w zakresie metodologii medycyny [Guyatt, Jaeschke, i współ. 2000]. Zmienia ona w zasadniczy sposób dotychczasowe ustalenia dotyczące pozyskiwania i oceny danych empirycznych dla procesu podejmowania decyzji klinicznych. Jednym z ważniejszych wyznaczników tej koncepcji jest formułowany w niej postulat hierarchizacji danych empirycznych.² Zgodnie z nim najbardziej wiarygodne są te dane, które pochodzą z randomizowanych badań kontrolnych (RCT), tzw. eksperymentów klinicznych.³ Randomizowane badania kon-

¹ Termin „evidence-based medicine” tłumaczony był w polskiej literaturze przedmiotu w różny sposób. Proponowane tłumaczenia budzą jednak wiele wątpliwości. Rozbieżność rozstrzygnięć translatorskich jest uciążliwa przede wszystkim dla potencjalnego czytelnika, dlatego też coraz częściej autorzy decydują się na posługiwanie terminem oryginalnym [Gajewski, 2003]. Takie rozwiązanie przyjęte jest również w niniejszym artykule.

² Artykuł nie ma na celu przedstawiania założeń i tez koncepcji EBM. Wskazany został wyłącznie ten aspekt tej koncepcji, który jest istotny z uwagi na prowadzone w dalszej części tekstu analizy. Współczesna literatura dotycząca tej koncepcji jest bardzo obszerna.

³ W artykule została przyjęta zasada posługiwanie się tymi tłumaczeniami terminów, które są najczęściej wykorzystywane w polskiej literaturze przedmiotu. Dotyczy to m.in. takich terminów jak: randomizacja, badanie kliniczne, badanie kliniczno-kontrolne, ślepa próba, badanie z randomizacją itp. Patrz np. tłumaczenie Z. Szawarskiego książki Wulffa i Göttsche’go [Wulff, Göttsche, 2005].

ontrolne są pewnym rodzajem badań klinicznych, a więc badań, w których ocenia się skuteczność terapii poprzez porównanie jej efektów w grupie osób poddanych terapii i grupie osób nieleczonych. W badaniu RCT przydział do odpowiedniej grupy ma charakter losowy. Zgodnie z koncepcją EBM wyniki badań RCT są najbardziej wiarygodne. Znacznie mniej wiarygodne są natomiast dane pochodzące z badań nieeksperymentalnych oraz dane uzyskiwane z modeli opisujących mechanizmy przyczynowe terapii w naukach podstawowych. W ciągu ostatnich dwóch dekad ugruntowane zostało przekonanie, że koncepcja EBM dostarcza jedynej prawidłowej oceny ewidencji empirycznej pozyskiwanej dla potrzeb decyzji terapeutycznych.

Postulat hierarchizacji ewidencji empirycznej pełni niezwykle istotną rolę we współczesnej praktyce medycznej. Będąc jednak silną tezą epistemologiczną, postulat ten wymaga odpowiedniego uzasadnienia [Worrall, 2002]. Tymczasem formułowana w literaturze przedmiotu argumentacja budzi wiele wątpliwości. Zasadnicze dotyczą wyróżnionego statusu danych uzyskiwanych z badań randomizowanych oraz związku pomiędzy wynikami tych badań a twierdzeniami opisującymi oddziaływanie przyczynowo-skutkowe zdarzeń klinicznych.

Zadaniem niniejszego artykułu jest realizacja dwóch celów. Pierwszym jest przedstawienie dyskusji nad funkcją randomizacji w procesie pozyskiwania ewidencji empirycznej dla potrzeb podejmowanych decyzji klinicznych. Drugim jest omówienie alternatywnej wobec RCT metody pozyskiwania ewidencji. W metodzie tej przeprowadza się analizę danych wykorzystującą teorię zbiorów przybliżonych. Zastosowanie tej metody pozwala ustalić związki przyczynowo-skutkowe zdarzeń klinicznych oraz określić algorytm decyzyjny postępowania lekarskiego.

2.

Pozyskiwanie danych empirycznych dla potrzeb podejmowanych decyzji klinicznych może przebiegać w różny sposób. Przez długi czas dominowało przekonanie, że decyzje o metodzie terapii powinny być uzasadniane teoriami i modelami nauk podstawowych opisującymi oddziaływanie terapeutyczne w kategoriach związków przyczynowo-skutkowych. W tym ujęciu decyzje terapeutyczne opierały się na wiedzy z zakresu biochemii, farmakokinetyki, patofizjologii oraz ustaleniach wielu innych dyscyplin szczegółowych. Posiadały one wyłącznie pośrednią ewidencję empiryczną: były potwierdzone, o ile potwierdzone empirycznie były teorie nauk podstawowych.

Bezpośrednia ewidencja empiryczna dotycząca efektywności danej terapii może być uzyskana w oparciu o odpowiednio zaprojektowane badania kliniczne. Wyróżnia się trzy główne ich typy: badania kohortowe, badania kliniczno-kontrolne oraz badania RCT. W badaniu kohortowym dokonuje się podziału osób poddanych badaniu na kilka grup określanych mianem kohort. W najprostszym przypadku są to dwie kohorty. Osoby należące do jednej z nich poddane zostają działaniu czynnika terapeutycznego.

tycznego. Z kolei osoby z drugiej kohorty tworzą tzw. grupę kontrolną. Zdarzenia zachodzące u osób poddanych terapii z pierwszej kohorty są następnie porównywane ze zdarzeniami występującymi u osób z grupy kontrolnej.⁴ Zaletą badania kohortowego jest to, że może być ono wykorzystane zarówno w celu oceny skuteczności terapii, jak i w celu oceny szkodliwości zdarzeń, które miały miejsce wcześniej.

Rozważmy krótki przykład. Załóżmy, że pragniemy określić, jaki jest wpływ palenia papierosów na występowanie nadciśnienia tętniczego. W badaniu kohortowym wyróżnia się dwie kohorty: kohortę osób palących i kohortę osób niepalących. Następnie ocenia się, jaki procent osób w każdej kohorcie cierpi na nadciśnienie tętnicze. Badanie tego typu jest badaniem o charakterze prospektywnym. Wnioskowanie przebiega od domniemanej przyczyny (palenie papierosów) do potencjalnych skutków (wystąpienie nadciśnienia tętniczego). W podobny sposób za pomocą badania kohortowego dokonuje się oceny efektywności terapii. Następuje podział pacjentów na dwie kohorty. W jednej z nich pacjenci otrzymują testowany lek, w drugiej natomiast nie są poddani leczeniu lub są leczeni uznaną metodą tradycyjną. Następnie ustala się, w której kohorcie jest większa liczba osób, u których wystąpiła poprawa stanu zdrowia. Wnioskowanie przebiega od domniemanej przyczyny (terapia) do potencjalnych skutków (poprawa stanu zdrowia).

W badaniu kliniczno-kontrolnym również dokonuje się porównania częstości występowania zdarzeń w uprzednio wyselekcjonowanych grupach. Jednak selekcja przebiega inaczej niż w badaniu kohortowym. Załóżmy, że pragniemy ustalić wpływ zdarzenia A na częstość występowania zdarzenia B. W badaniu kliniczno-kontrolnym podział na kohorty zostaje dokonany z uwagi na występowanie lub brak zdarzenia B, które jest przypuszczalnym następstwem zdarzenia A. W nawiązaniu do poprzedniego przykładu, byłby to zatem podział na kohortę osób, u których stwierdza się nadciśnienie tętnicze i kohortę osób u których nadciśnienie tętnicze nie występuje. Następnie bada się częstość występowania w obu kohortach domniemanego czynnika przyczynowego A. Dla rozważanego przykładu ustala się zatem, jaka liczba osób z każdej kohorty pali papierosy. Badanie tego typu jest badaniem retrospektywnym. Wnioskowanie przebiega od rozważanego skutku (nadciśnienie lub jego brak) do domniemanej przyczyny (palenia papierosów). Podobnie przebiega badanie kliniczno-kontrolne w przypadku oceny skuteczności terapii. Podział na kohorty zostaje dokonany na podstawie tego, czy pacjent wyzdrowiał, czy nie. Następnie analizuje się, jaka terapia została zastosowana w przypadku kohorty osób, które wyzdrowiały, a jaka w przypadku osób, u których nie stwierdzono poprawy stanu zdrowia.

⁴ Przy większej ilości kohort można porównywać kilka różnych typów leczenia. Warto przy tym zauważyć, że w przypadku uzyskania wyników wskazujących na wysoką skuteczność w leczeniu danego schorzenia, kolejne potencjalne badania kliniczne nie mogą już posiadać „czystej” grupy kontrolnej. Nieetyczne byłoby bowiem pozostawienie grupy pacjentów bez terapii, w sytuacji, w której posiadamy wiedzę o tym, że istnieje terapia skuteczna. Porównuje się wówczas wyłącznie efekty nowej techniki terapeutycznej (wobec, której lekarze muszą posiadać uzasadnione przekonanie (!), że jest potencjalnie skuteczna) z terapią, której skuteczność została ustalona już wcześniej.

Wnioskowanie zatem również przebiega od rozważanego skutku (poprawa stanu zdrowia) do ewentualnej przyczyny (zastosowanie danej terapii).⁵

Randomizowane badanie kontrolne jest odmianą badania kohortowego. W badaniu tym następuje podział na kohorty osób poddanych leczeniu lub nieleczonych. Jest to zatem badanie, w którym wnioskowanie również przebiega od potencjalnej przyczyny (terapia) do domniemanego skutku (poprawa stanu zdrowia). W przeciwieństwie do badania kohortowego w badaniu RCT przydział do poszczególnych kohort ma jednak charakter losowy. Randomizacja jest jedynym czynnikiem różniącym to badanie od badania kohortowego.⁶ Badanie RCT może być badaniem z podwójnie ślełą próbą (zarówno pacjent, jak i lekarz nie wiedzą czy zastosowana została terapia, czy placebo) lub badaniem z pojedynczo ślełą próbą (albo pacjent, albo lekarz nie wie, czy zastosowano terapię, czy placebo). Uważa się, że procedura randomizacji zapewnia uzyskiwanie najbardziej wiarygodnych informacji o skuteczności podjętych działań terapeutycznych.

Postulat przeprowadzania randomizacji w procesie pozyskiwania danych empirycznych po raz pierwszy sformułowany został w 1935 r. przez R. Fishera.⁷ Praca Fishera dotyczyła oceny stosowania odmiennych metod upraw w rolnictwie. Pierwsze badanie RCT w medycynie zostało wykonane w 1948 r. w celu oceny skuteczności stosowania streptomycyny w leczeniu gruźlicy. Z kolei w 1950 r. po raz pierwszy w eksperymencie randomizowanym z podwójnie ślełą próbą zastosowano placebo [Brzeziński, Szamotulska, 1997]. Rozwój koncepcji EBM rozpowszechnił przekonanie o tym, że badania z randomizacją dostarczają najbardziej wiarygodnej ewidencji empirycznej ustalającej skuteczność terapii. Jak pisał Straus:

Jeżeli artykuł dostarcza danych z badań nierandomizowanych, to najlepiej od razu przerwać jego czytanie i szukać badań RCT [Straus, 2005, s. 118].

Badania nierandomizowane powinny stanowić podstawę dla podejmowanych decyzji klinicznych wyłącznie wówczas, gdy nie ma przeprowadzonych badań RCT. Te ostatnie określane są współcześnie mianem „złotego standardu” w metodologii medycyny. Jakże jednak argumenty przemawiają na rzecz tezy, że randomizacja dostarcza najbardziej wiarygodnych danych empirycznych dla potrzeb podejmowanych decyzji klinicznych? Zdaniem niektórych filozofów nauki, pytanie to identyfikuje

⁵ Badania kliniczno-kontrolne często przeprowadzane są w celu określenia niekorzystnych efektów ubocznych terapii. Ponadto jest to badanie, które przeprowadza się dla ustalenia odległych czasowo punktów pomiaru, przykładowo, jaki był efekt długoterminowy terapii na przeżycie pacjentów po kilku, lub kilkunastu latach od zakończenia terapii (często przeprowadzane w onkologii i kardiologii).

⁶ Jest to jednak czynnik niezwykle istotny, chociażby z uwagi na fakt, że wyklucza jako nietyczne próby ustalenia wpływu czynników szkodliwych na zdrowie pacjenta. Przykładowo, nie można ustalić za pomocą tego badania, czy palenie jest czy też nie jest szkodliwe.

⁷ Początków postulatu randomizacji można doszukiwać się wśród prac poświęconych eksperymentom z zakresu parapsychologii i zjawisk parapsychicznych [Hacking, 1988].

jeden z najbardziej interesujących współcześnie problemów epistemologicznych [Worrall, 2002].

3.

Ocena skuteczności terapii w medycynie wymaga ustalenia, czy oddziaływanie rozważane jako rodzaj terapii stanowić będzie przyczynę oczekiwanych efektów. Załóżmy, że badacz dysponuje układem fizycznym A (pacjent) oraz rozważa terapię T jako potencjalną przyczynę uzyskania oczekiwanego efektu R w układzie A . Problem polega na tym, że uzyskanie efektu R , po zastosowaniu terapii T wobec układu A , nie przesądza, że to właśnie T jest przyczyną wystąpienia efektu R . Pragnąc taki wniosek sformułować, należałoby porównać układ A poddany działaniu czynnika T z identycznym układem A' niepoddanym działaniu tego czynnika. A' stanowi wówczas jednoelementową grupę kontrolną. Eksperyment, w którym badacz dysponuje dwoma identycznymi układami fizycznymi różniącymi się wyłącznie jednym czynnikiem stanowiącym przedmiot badań, Fisher określa mianem eksperymentu idealnego [Urbach, 1985]. Niemniej Fisher zdawał sobie sprawę, że tego rodzaju eksperyment jest niewykonalny, ponieważ rozważane układy fizyczne (poletka doświadczalne, pacjenci) zawsze będą różniły się bardzo wielu czynnikami, które nie są badaczowi znane, a które mogą posiadać wpływ na występowanie lub brak oczekiwanego efektu. Czynniki takie zwane są czynnikami zakłócającymi. Ich występowanie w badanym układzie może sprawiać systematyczne wypaczanie obserwowanych wyników [Jaeschke, Cook, Guyatt, 1998a]. Fisher miał świadomość tego problemu. Dlatego też postulował przeprowadzanie randomizacji, która — w jego przekonaniu — w największym stopniu pozwala zrealizować cel eksperymentu idealnego. Po pierwsze, proces randomizacji gwarantuje, zdaniem Fishera, równomierny rozkład czynników zakłócających [Urbach, 1985, s. 257]. Po drugie, randomizacja zwiększa efektywność eksperymentu w tym sensie, że pozwala uzyskać statystycznie istotne różnice efektów przy niewielkiej ilości badanych obiektów. Drugi z tych argumentów określany jest współcześnie mianem argumentu z istotności statystycznej [Worrall, 2002].

Rozwój koncepcji EBM przyczynił się do rozpowszechnienia argumentacji na rzecz tezy o wyróżnionym statusie epistemicznym badań z randomizacją. Można wskazać na dwa najczęściej przytaczane argumenty na rzecz tej tezy. Po pierwsze, uznaje się, że randomizacja ogranicza wpływ czynników subiektywnych (tj. przekonań i oczekiwań) na przebieg badania. Randomizacja umożliwia bowiem utajnienie przydziału pacjenta do grupy kontrolnej lub do grupy osób poddanych leczeniu poprzez zastosowanie ślepej próby. Wskazać można w tym przypadku na trzy zasadnicze aspekty ograniczenia wpływu czynników subiektywnych na wynik badania. Pierwszym jest przydział pacjenta do grupy. Zastosowanie ślepej próby gwarantuje, że wiedza lekarza o stanie pacjenta nie będzie wpływała na jego alokację. Stwierdzono, że w badaniach, w których lekarz świadomie przydzielał pacjenta uzyskiwa-

no bardziej optymistyczne dane o skuteczności terapii [Jaeschke, Cook, Guyatt, 1998a, s. 152]. Drugi aspekt dotyczy oceny stanu pacjenta po przeprowadzonym badaniu. Zastosowanie ślepej próby gwarantuje w tym przypadku, że wiedza lekarza dokonującego oceny stanu pacjenta nie będzie miała wpływu na tę ocenę. Problem ten dotyczy przede wszystkim tych badań, w których przyjęto tzw. miękkie punkty końcowe, których ocena ma wyraźnie subiektywny charakter (np. ocena wyników dotyczyła zmian radiologicznych). Trzeci aspekt związany jest z subiektywnymi odczuciami pacjenta dotyczącymi efektów terapii. Wiedza pacjenta o tym, do jakiej grupy należy w oczywisty sposób może wpływać na formułowane przez niego sprostowania dotyczące efektów terapii lub ich braku.

Przytoczony argument na rzecz procedury randomizacji może budzić jednak wątpliwości. Okazuje się bowiem, że to nie losowy przydział pacjentów, lecz zastosowanie ślepej próby umożliwia pozyskanie danych empirycznych o wyróżnionym statusie epistemicznym [Abel, Koch, 1999, s. 491]. Randomizacja jest w tej sytuacji wyłącznie warunkiem umożliwiającym przeprowadzenie ślepej próby. Niektórzy autorzy podkreślają nawet, że nie jest ona konieczna. Można bowiem dokonać utajnienia przydziału w inny sposób. Akceptując stanowisko bayesianizmu, Urbach utrzymuje, że

...nie ma żadnych korzyści, które wynikałyby z przydziału pacjentów do grupy kontrolnej lub testowej na podstawie procedury randomizacji. Dowolna metoda przydziału byłaby równie satysfakcjonująca pod warunkiem, że nie mielibyśmy powodów, aby sądzić, że posiada ona wpływ na wyniki przeprowadzanego badania [Urbach, 1985, s. 270].⁸

Polemika z tą tezą zostanie podjęta w dalszej części artykułu.

Drugi argument zwolenników koncepcji EBM na rzecz randomizacji nawiązuje bezpośrednio do tezy sformułowanej przez Fishera. Dotyczy ona kwestii umniejszenia wpływu czynników zakłócających wyniki badań. Realizacja tego celu wymaga, aby rozkład czynników zakłócających był taki sam w grupie osób poddanych leczeniu, jak i w grupie kontrolnej. Powtarzając tezę Fishera, zwolennicy EBM utrzymują, że taki równomierny rozkład zapewnia procedura randomizacji [Jaeschke, Cook, Guyatt, 1998a, s. 151]. Można odnieść wrażenie, że teza ta nie budzi współcześnie żadnych wątpliwości. Tak jednak nie jest. Stanowi ona przedmiot krytyki formułowanej zarówno przez filozofów nauki, statystyków, jak i samych klinicystów. Warto więc przybliżyć kilka najważniejszych punktów tej krytyki.

4.

Oceniając ewidencję empiryczną uzyskaną z badań RCT, lekarz musi ustalić, czy analizowane badanie zostało przeprowadzone we właściwy sposób. W szczególno-

⁸ W podobny sposób Worrall rekonstruuje stanowisko Savage'a wobec procedury randomizacji [Worrall, 2007, s. 482].

ści: czy grupy porównywane w badaniu RCT miały podobne charakterystyki na początku badania, czyli czy randomizacja była skuteczna [Jaeschke, Cook, Guyatt, 1998a, s. 151]? Rozstrzygnięcie tej kwestii może być przeprowadzone wyłącznie w odniesieniu do pewnej skończonej liczby wybranych czynników. Ponadto obie grupy mogą jednak różnić się ze względu na występowanie wielu czynników niebranych pod uwagę. Oznacza to, że uzyskanie równomiernego rozkładu znanych czynników zakłócających można traktować wyłącznie jako zdarzenie zwiększające prawdopodobieństwo równomiernego rozkładu czynników nieznanych. Nie możemy jednak mieć **pewności**, że wszystkie nieznanne czynniki występują w obu grupach w zbliżonych proporcjach [Worrall, 2002], [Worrall, 2007], [Urbach, 1985]. Przyjrzyjmy się bliżej tej kwestii.

Założmy, że przeprowadzono procedurę randomizacji, w wyniku której pacjenci zostali przydzieleni do dwóch kohort. Przyjmijmy jednocześnie, że dostępna wiedza pozwala sądzić, że czynniki f_1, f_2, f_3 mogą posiadać wpływ na przebieg terapii. Pragnąc zatem sprawdzić, czy procedura randomizacji spełniła nasze oczekiwania, jeżeli chodzi o rozkład czynników zakłócających, podejmiemy próbę ustalenia rozkładu czynników $\{f_1, f_2, f_3\}$ w obu grupach. Po ustaleniu rozkładu tych czynników w obu grupach o czynnikach tych powiemy, że są to czynniki znane. Pozostałe czynniki określimy mianem nieznanych. Przeprowadzona randomizacja podlega zatem ocenie zgodnie z następującą regułą:

N: Procedura randomizacji jest wykonana prawidłowo o tyle, o ile wystąpił równomierny rozkład znanych czynników w obu grupach.⁹

Niech $P(R_{CN}/R_{CZ})$ oznacza prawdopodobieństwo równomiernego rozkładu czynników nieznanych ze względu na równomierny rozkład czynników znanych, a $P(R_{CN})$ to prawdopodobieństwo a priori uzyskania równomiernego rozkładu czynników nieznanych. Wówczas wnioskowanie zwolenników randomizacji można zrekonstruować następująco:

- (1) Jeżeli w porównywanych grupach występuje równomierny rozkład znanych czynników, to $P(R_{CN}/R_{CZ}) > P(R_{CN})$.
- (2) W porównywanych grupach występuje równomierny rozkład znanych czynników — (na podstawie oceny wiarygodności badania RCT).
- (3) Zatem $P(R_{CN}/R_{CZ}) > P(R_{CN})$.

Wnioskowanie to — jakkolwiek prawidłowe (modus ponendo ponens) — nie uzasadnia jednak tezy, że to właśnie randomizacja jest procedurą gwarantującą uzyskiwa-

⁹ Do tej kwestii odnosił się również Fisher. Utrzymywał, że w przypadku przeprowadzania randomizacji eksperymentator powinien sprawdzić rozkład znanych czynników. Jeżeli rozkład tych czynników nie jest równomierny, to procedurę randomizacji należy powtarzać tak długo, aż rozkład będzie równomierny [Urbach, s. 260, 1985].

nie wiarygodnej ewidencji empirycznej. Jeżeli równomierny rozkład znanych czynników (R_{CZ}) zwiększa prawdopodobieństwo uzyskania równomiernego rozkładu czynników nieznanymi, to R_{CZ} może również stanowić kryterium przydziału pacjentów do porównywanych grup. W tej sytuacji procedura randomizacji staje się jednak zbędna, ponieważ akceptacja kryterium R_{CZ} eliminuje potrzebę jej przeprowadzania.¹⁰

Zwolennicy badań z randomizacją mogą jednak utrzymywać, że należy odróżnić kryterium oceny randomizacji od kryterium alokacji pacjentów. Ocena randomizacji może być przeprowadzona na podstawie pewnego skończonego zbioru czynników. Czynniki te mogą jednak nie wystarczać do przeprowadzenia alokacji pacjentów. Rozważmy dwa przypadki. W pierwszym, skończony zbiór czynników $\{f_1, \dots, f_n\}$ stanowi podstawę oceny randomizacji. Równomierny rozkład tych czynników będzie zatem świadczył o prawidłowym przebiegu tej procedury.

W drugim przypadku zbiór czynników $\{f_1, \dots, f_n\}$ stanowi kryterium przydzielania pacjentów do porównywanych grup. Przyjmijmy, że osoba dokonująca przydziału pacjentów nie posiada wykształcenia medycznego. Zgodnie z ustalonym wskazaniem osoba ta ma uzyskać porównywalny rozkład czynników f_1, \dots, f_n w obu grupach: testowej i kontrolnej. Jednakże alokacja pacjentów, spełniająca ustalone wskazanie, może być przeprowadzona w różny sposób. Osoba dokonująca przydziału może kierować się w swoim wyborze dowolną cechą φ_i różnicującą zbiór pacjentów (kolor oczu, wzrost itp.). Do grupy testowej zostaną wówczas przydzieleni pacjenci, dla których „wartość” cechy φ_i będzie mieściła się w pewnym przyjętym arbitralnie przedziale (np. niebieski kolor oczu lub wzrost powyżej 175 cm itp.). Do grupy kontrolnej natomiast zostaną przydzieleni pozostali pacjenci. Oznacza to jednak, że osoba dokonująca przydziału musiała zaakceptować dodatkowo pewną regułę alokacji G . Reguła ta postulowała przeprowadzenie podziału logicznego uniwersum pacjentów ze względu na „wartości” cechy φ_i . Zasadą podziału była relacja równoważnościowa wyznaczająca klasy abstrakcji identyczne z członami podziału, czyli grupą testową i grupą kontrolną. Taki przydział pacjentów do obu grup budzi jednak zastrzeżenia zwolennika randomizacji. Cecha φ_i może bowiem być wielkością skorelowaną statystycznie z nieznanymi nam czynnikami, które posiadają wpływ na wyniki przeprowadzanej terapii. W takim przypadku akceptacja reguły alokacji może prowadzić do systematycznego wypaczania wyników badania, pomimo że przydział pacjentów był przeprowadzany przez podmiot nieposiadający wiedzy medycznej. W efekcie podmiot kierujący się regułą alokacji postulującą podział logiczny przeprowadzony ze względu na cechę φ_i zwiększa prawdopodobieństwo nierównomiernego rozkładu czynników nieznanymi. Na ten fakt zwraca uwagę N. Cartwright, pisząc:

¹⁰ Podobny wniosek sformułowany został przez P. Urbacha [Urbach, 1985, s. 271]. Według Worralla przedstawione rozumowanie można przypisać większości zwolennikom bayesianizmu [Worrall, 2007, s. 463].

przydział pacjenta do grupy testowej lub kontrolnej powinien być statystycznie niezależny od wszystkich cech pacjenta, które mogą być przyczynowo istotne dla wyniku badania [Cartwright, 1989, s. 64].

Nietrafna wydaje się zatem argumentacja Urbacha utrzymującego, że przydział pacjentów może być dokonany przez osobę nieposiadającą wiedzy medycznej. Podmiot dokonujący przydziału pacjentów powinien bowiem ustalić regułę alokacji wobec tych cech, które nie są w żaden sposób skorelowane z cechami biologicznymi pacjentów. Może to być jednak zadanie trudne, ponieważ nie dysponując wiedzą medyczną podmiot nie jest w stanie ustalić, która z wybranych przez niego cech takiej korelacji nie wykazuje. Zadanie to będzie równie trudne dla osoby posiadającej wiedzę medyczną. Współcześnie wiadomo, że takie czynniki jak rodzaj pracy, miejsce zamieszkania, sytuacja rodzinna mogą posiadać wpływ na wyniki terapii, gdyż są skorelowane z cechami biologicznymi. Nie wiemy jednak, jakie inne czynniki mogą mieć również wpływ na wyniki badań. W tej sytuacji najbardziej bezpieczną metodą przydziału jest zastosowanie procedury randomizacji. Stanowi ona szczególny rodzaj reguły alokacji wyznaczonej zbiorem liczb losowych, dziecięcą wyliczanką lub rzutem kostką. W każdym z tych przypadków reguła alokacji jest ustalona ze względu na cechę niemającą żadnego związku z pacjentem.

Zastosowanie procedury randomizacji nie gwarantuje oczywiście, że nastąpi równomierny rozkład czynników nieznanych [Worrall, 2002], [Urbach, 1985]. Procedura ta wyklucza jednak sytuację, w której podmiot dokonujący przydziału pacjentów przyjmuje (świadomie lub nieświadomie) regułę alokacji postulującą podział logiczny pacjentów ze względu na taką cechę, która może być skorelowana statystycznie z cechami biologicznymi pacjentów. Okazuje się zatem, że zasadniczą funkcją randomizacji jest wyeliminowanie wpływu czynników subiektywnych na przydział pacjentów. Oznacza to jednak, że najistotniejszy argument zwolenników procedury randomizacji redukuje się do przytoczonego wcześniej argumentu pierwszego: randomizacja pozwala wykluczyć wpływ czynników subiektywnych na przebieg badania.¹¹

5.

Omówione argumenty na rzecz randomizacji nie wyczerpują jednak strategii przyjmowanych przez zwolenników tej procedury. Warto wspomnieć o filozoficznej argumentacji formułowanej na gruncie probabilistycznych koncepcji przyczynowości. Podstawowa idea tych koncepcji sprowadza się do stwierdzenia, że zależności o charakterze probabilistycznym mogą być wyjaśniane w kategoriach związków przyczynowych [Cartwright, 2010, s. 60]. Autorzy tych koncepcji starają się wykazać, że procedura randomizacji pozwala uzasadnić twierdzenia ustalające związki przyczynowo-skutkowe o efektywności stosowanych terapii. Takie stanowisko zaj-

¹¹ Na fakt tego rodzaju redukcji argumentu zwracał również uwagę Worrall [Worrall, 2007].

muje między innymi D. Papineau [Papineau, 1994]. Jego zdaniem, procedura randomizacji zapewnia, że zastosowana terapia nie jest skorelowana z żadnymi czynnikami zakłócającymi, które posiadałyby wpływ na efekty badania. W konsekwencji, jeżeli w badaniu z randomizacją stwierdzamy, że:

$$P(R/T) > P(R/\neg T),$$

to możemy wnosić, że T jest przyczyną wystąpienia zdarzenia R [Papineau, 1994, s. 440]. Podejmując polemikę z tą tezą, Worrall proponuje rozważenie przykładu opartego na następujących założeniach: (i) w przypadku choroby D zastosowanie terapii T prowadzi do wystąpienia zdarzenia R (wyzdrowienia), o ile u pacjenta występują czynniki f_1, \dots, f_i , (ii) czynniki f_1, \dots, f_i bardzo rzadko występują w populacji i nie są znane lekarzom, (iii) w przypadku braku tych czynników terapia T jest przyczyną wystąpienia niekorzystnych efektów ubocznych u danego pacjenta np. wzmożenia objawów chorobowych $\neg R$. Załóżmy, że (iv) przeprowadzono badanie z randomizacją i że procedura randomizacji doprowadziła do równomiernego rozkładu wszystkich czynników znanych i nieznanymi w grupie testowej i kontrolnej. Jednakże ze względu na fakt, że czynniki f_1, \dots, f_i bardzo rzadko występują w populacji, stwierdzono, że:

$$g(R/T) < g(R/\neg T), \text{ gdzie } g(R/T) \text{ lub } g(R/\neg T) \text{ to względna częstość zdarzenia } R \text{ odpowiednio ze względu na } T \text{ lub } \neg T.$$

Badacze powinni zatem wnosić, że terapia T jest przyczyną $\neg R$ [Worrall, 2007, s. 467]. Wniosek ten jest jednak błędny z uwagi na założenie (i). Worrall przyznaje, że rozważany przykład jest fikcyjny, niemniej pozwala on wykazać, że randomizacja nie może stanowić podstawy dla wniosków ustalających zależności przyczynowe. W dalszej części artykułu pokażemy, że przykład ten nie odbiega od przypadków mających miejsce w rzeczywistej praktyce medycznej. Pozwoli on natomiast dokonać istotnych — z perspektywy dalszych rozważań — spostrzeżeń.

S_I : Ustalenie związku przyczynowego pomiędzy zdarzeniami T i R wymaga wskazania czynników f_1, \dots, f_i , których wystąpienie razem z T przesądza o wystąpieniu R .¹²

Na podstawie S_I można sformułować charakterystykę warunkowego związku przyczynowego.

C : Zdarzenie T jest przyczyną zajścia zdarzenia R , o ile w rozważanym układzie wystąpiły dodatkowo czynniki f_1, \dots, f_i .

Zbliżoną charakterystykę związku przyczynowego akceptuje N. Cartwright. Formuluje ona następujące kryterium dla ustalenia zależności przyczynowych.

¹² Spostrzeżenie to jest zgodne z przyczynowym ujęciem procedury wyjaśniania proponowanym w koncepcji wyjaśniania przez wyszczególnienie [Kuipers, Wiśniewski, 1994].

CC: Zdarzenie T jest przyczyną zdarzenia R wtw $P(R/T \wedge K) > P(R/\neg T \wedge K)$,

gdzie K jest zbiorem wszystkich czynników f_1, \dots, f_i przyczynowo istotnych dla wystąpienia zdarzenia R (oczywiście poza zdarzeniem T [Cartwright, 1989, s. 79], [Cartwright, 2010, s. 60]). Przyczynowa istotność czynników f_1, \dots, f_i oznacza, że są to czynniki, których wystąpienie jest warunkiem koniecznym, tego, aby wystąpienie zdarzenia T doprowadziło do zajścia zdarzenia R . Czynniki te nie są natomiast wystarczające dla wystąpienia zdarzenia R .

Dokonane przez Cartwright wyróżnienie zbioru K jest ważne z uwagi na problem ekstrapolowania wyników badań klinicznych.¹³ Problem ten można scharakteryzować następująco. Przyjmijmy, że w efekcie przeprowadzonego badania klinicznego stwierdzono, że pomiędzy zdarzeniami R i T zachodzi korelacja statystyczna k : $P(R/T) > P(R/\neg T)$. Czy można mieć jednak pewność, że zależność k będzie również obowiązywała w docelowej populacji pacjentów, u których rozpoznano schorzenie D ? Badanie kliniczne nie dostarcza tego rodzaju gwarancji, ponieważ nie wiemy, czy grupa osób zakwalifikowana do badań stanowi grupę reprezentatywną dla całej populacji pacjentów [Worrall, 2009, s. 296]. Kryterium CC sformułowane przez Cartwright pozwala ten problem rozwiązać. Jeżeli bowiem czynniki f_1, \dots, f_i zostały rozpoznane jako wszystkie czynniki przyczynowo istotne, to można wnosić, że korelacja statystyczna ustalona w badaniu klinicznym będzie również obowiązywać w każdej docelowej grupie pacjentów, u których czynniki te występują. Grupa docelowa jest wówczas przyczynowo homogeniczna z grupą osób, które brały udział w badaniu klinicznym [Worrall, 2007, s. 469], [Cartwright, Munro, 2010, s. 261]. Ustalenia te pozwalają sformułować kolejne spostrzeżenie.

S_2 : Zdarzenie T będzie przyczyną wystąpienia zdarzenia R w populacji docelowej A wtw, T było przyczyną zajścia zdarzenia R w populacji B oraz B i A są przyczynowo homogeniczne z uwagi na zbiór czynników K .¹⁴

O tym, czy zdarzenie T było przyczyną zajścia zdarzenia R w populacji B (pacjentów biorących udział w badaniu klinicznym) można natomiast, zdaniem Cartwright, rozstrzygnąć na podstawie kryterium CC. Jednakże kryterium CC nie zapewnia, że w porównywanych grupach wystąpi równomierny rozkład czynników istotnych przyczynowo. W tym miejscu swoich rozważań Cartwright wprowadza pojęcie *idealnego badania RCT*. Idealne badanie RCT charakteryzuje jako badanie, w którym zastosowanie procedury randomizacji powoduje równomierny rozkład czynników ze zbioru K . Wobec idealnego badania RCT nie znajdowałby zatem za-

¹³ Zagadnienie to określane jest w literaturze przedmiotu terminem „external validity” lub „transferability problem” [Cartwright, 2010], [Worrall, 2007], [Roush, 2010]. W polskiej literaturze przedmiotu termin ten jest również tłumaczony jako „wiarygodność zewnętrzna” [Jaeschke, Cook, Guyatt, 1998b, s. 115].

¹⁴ S_2 jest uproszczoną wersją zasady ustalającej poprawność ekstrapolacji danych podanej przez Cartwright [Cartwright, 2010, s. 66].

stosowania argument sformułowany przez Worralla na podstawie jego „fikcyjnego” przykładu. Zbiór K jest bowiem na mocy definicji zbiorem zawierającym wszystkie przyczynowo istotne czynniki. Pozostałe cechy charakteryzujące pacjentów są przyczynowo nieistotne. Problem polega jednak na tym, że — jak zaznacza sama Cartwright — nie wiemy, w jakim stopniu przeprowadzona randomizacja jest bliska idealnemu badaniu RCT [Cartwright, 2010, s. 64]. Z drugiej strony warto zauważyć, że gdybyśmy wiedzieli, które czynniki są przyczynowo istotne dla wyników badania, to randomizacja byłaby procedurą zbędną. Można by bowiem dokonać przydziału pacjentów do grup (testowej lub kontrolnej), kierując się równomiernym rozkładem czynników ze zbioru K oraz dowolnie ustaloną regułą alokacji G dotyczącą czynników, które nie są skorelowane z czynnikami przyczynowo istotnymi dla wyników badania.¹⁵ Postulat rozważania idealnego badania RCT jest zatem mało przekonujący. Na większą uwagę zasługują natomiast ustalenia Cartwright dotyczące konieczności wyróżnienia zbioru K . Spostrzeżenie S_2 pozwala bowiem sformułować ogólną wersję algorytmu decyzyjnego.

DAD: Niech K będzie zbiorem wszystkich czynników przyczynowo istotnych dla wystąpienia zdarzenia R po zastosowaniu terapii T . Wówczas:

Dla dowolnej populacji A , charakteryzującej się zbiorem cech α , takich, że $K \subseteq \alpha$, obowiązuje zależność $T(A) \rightarrow R(A)$.

Czyli, zastosowanie terapii T w stosunku do populacji A jest warunkiem wystarczającym zajścia R w owej populacji.

Zauważmy, że *DAD* jest algorytmem decyzyjnym o charakterze deterministycznym. Sformułowanie takiego algorytmu dla konkretnych przypadków klinicznych pozwoliłoby na jednoznaczne ustalenie schematów podejmowania decyzji terapeutycznych.

6.

Rozważmy przykład zbliżony do przykładu sformułowanego przez Worralla w jego polemice z Papineau. Załóżmy, że terapię T zastosowano wobec osób, u których rozpoznano chorobę D . Stwierdzono, że zastosowanie tej terapii wobec niektórych pacjentów doprowadziło do wystąpienia korzystnego efektu R , podczas gdy u innych ta sama terapia spowodowała efekt niekorzystny $\neg R$. Ze względu na niezgodność uzyskanych wyników można wnosić, że na efekt końcowy terapii wpływ posiadały pewne dodatkowe czynniki zakłócające. W tym przypadku przeprowadzenie badania RCT doprowadziłoby wyłącznie do „uśrednienia” niezgodnych wyników. Przykład ten wskazuje na istotne ograniczenie zastosowania badań RCT. Wbrew zastrzeżeniom Worralla tego rodzaju „fikcyjne” przykłady nie odbiegają od

¹⁵ Nie byłoby wówczas powodu do formułowania zastrzeżeń wskazanych w paragrafie 4.

przypadków zachodzących w „rzeczywistej” praktyce medycznej. W dalszej części artykułu odwołam się do przykładu oceny skuteczności operacyjnego leczenia choroby wrzodowej dwunastnicy.

Od lat 70. XX wieku jedną z najczęściej stosowanych metod leczenia choroby wrzodowej dwunastnicy była *wagotomia wysoce wybiórcza*. Technika ta polega na selektywnym przecięciu nerwów odpowiedzialnych za nadmierne wydzielanie soku żołądkowego. Spośród różnych technik operacyjnych ta metoda prowadziła do najmniejszych uszkodzeń organów, pozwalając zachować ich funkcję, gwarantując zarazem utrzymanie określonego komfortu życia. Jednak chirurdzy nie byli zgodni w ocenie długoterminowych efektów tej terapii [Słowiński, 1990, s. 5]. W literaturze przedmiotu opisywane były zarówno przypadki pacjentów, u których stwierdzono poprawę po przeprowadzonej terapii, jak i pacjentów, którym terapia ta nie przyniosła żadnych korzyści. Uzyskiwanie niezgodnych wyników wskazywało na wpływ czynników zakłócających przebieg terapii.

Zauważmy, że czynniki zakłócające stanowią w istocie cechy charakteryzujące pacjenta. Niektóre z nich są cechami ściśle związanymi z przebiegiem i stopniem rozwoju choroby. W omawianym przypadku były to takie cechy jak: czas trwania choroby, powikłania wrzodu, zalegająca objętość soku żołądkowego, stężenie HCL itp. Cechy te stanowiły zbiór wskaźników diagnostycznych. Różne wartości tych wskaźników identyfikować mogły odmienne stany choroby *D*. Ponadto brano pod uwagę również takie cechy, których wartości nie zależały od przebiegu choroby (np. wiek i płeć), ale z kolei przebieg choroby (oraz oczywiście efekt terapii) mógł zależeć od wartości tych cech. Wszystkie te cechy określić można mianem atrybutów. Uzyskiwanie odmiennych wyników terapii świadczyło o tym, że różne wartości tych atrybutów posiadają wpływ na efekty leczenia. Jednak chirurdzy nie byli zgodni, nie tylko w odniesieniu do tego jakie wartości, ale również jakie atrybuty posiadają zasadniczy wpływ na wyniki terapii. Odmienne precyzowano wskazania do wykonywania operacji. Niektórzy autorzy uznawali, że najistotniejszym przyczynowo atrybutem jest czas trwania choroby. Inni jako podstawę do wykonywania zabiegu przyjmowali wiek pacjenta. Najbardziej niezgodne opinie ujawniały się jednak w odniesieniu do wartości atrybutów związanych z wydzielaniem żołądkowym. Jako wskazanie do wykonywania zabiegu różni autorzy uznawali skrajne wartości tych atrybutów [Słowiński, 1990, s. 47].

Opisana sytuacja identyfikuje ważny problem epistemologiczny. Jaki sposób pozyskiwania danych empirycznych pozwoliłby na przeprowadzenie wiarygodnej oceny efektywności stosowanej techniki operacyjnej? Badania kliniczne dokonywały wyłącznie uśrednienia uzyskiwanych wyników i w żaden sposób nie odzwierciedlały zróżnicowanych efektów uzyskiwanych w praktyce klinicznej [Słowiński, 1990, s. 5-6]. Pragnąc określić wskazania do podjęcia terapii, należało ustalić, które z rozważanych atrybutów są przyczynowo istotne oraz jakie ich wartości przesądzą o efektach terapii. W celu określenia takiego algorytmu decyzyjnego zastosowano metodę analizy danych wykorzystującą teorię zbiorów przybliżonych [Rough Set Theory —

RST]. W dalszej części artykułu w sposób nieformalny zostaną przedstawione zasadnicze idee związane z zastosowaniem tej teorii w medycynie dla potrzeb analizy danych empirycznych.¹⁶

7.

Teoria zbiorów przybliżonych sformułowana została w latach 80. XX wieku przez Z. Pawlaka. Wykorzystanie tej teorii przez K. Słowińskiego dla potrzeb analizy wyników wagotomii wysoce wybiórczej było jej pierwszym zastosowaniem w medycynie [Pawlak, Słowiński, 1986].

Przyjmijmy następujące wstępne ustalenia. Załóżmy, że $\Omega = \{Y_1, \dots, Y_k\}$ jest klasyfikacją pacjentów poddanych terapii T z uwagi na uzyskane efekty: R_1, \dots, R_j . Elementy rodziny Ω określimy mianem klas wyleczenia. Zauważmy, że efekt terapii jest również atrybutem, jakkolwiek nie jest on elementem zbioru K . Jest to cecha charakteryzująca pacjenta po przebytej terapii. Ustalenie algorytmu decyzyjnego (czyli dla rozważanego przykładu: sprecyzowanie wskazań i przeciwwskazań do operacji) wymaga określenia takiego zakresu wartości atrybutów ze zbioru K , że zastosowanie terapii T przy określonych wartościach przyjmowanych przez f_1, \dots, f_i jest przyczyną wystąpienia zdarzenia R_n , dla $1 \leq n \leq j$.

Należy zauważyć, że wiele atrybutów charakteryzujących stan pacjenta, to atrybuty, którym przypisywane są wartości ze zbioru liczb rzeczywistych. Przyporządkowanie wartości atrybutów do poszczególnych elementów rodziny Ω wymaga zatem przekształcenia wartości ilościowych na wartości jakościowe.¹⁷ Przekształcenie takie ma charakter arbitralny. Opiera się ono na ustalonych dla potrzeb diagnostyki wartościach referencyjnych i związane jest z charakterystyką testów diagnostycznych. Zagadnienia te nie będą w niniejszym artykule przedstawiane.

Problem oceny techniki wagotomii wysoce wybiórczej można uogólnić do następującej postaci. Załóżmy, że: (i) rozważamy zbiór pacjentów $U = \{a_1, \dots, a_8\}$, u których rozpoznano chorobę D , (ii) Zbiór $K = \{f_1, f_2, f_3\}$, gdzie f_1, f_2, f_3 to atrybuty charakteryzujące pacjenta przed zastosowaną terapią T , (iii) atrybuty f_1 oraz f_2 przyjmują wartości 0 lub 1, (iv) atrybut f_3 może przyjmować wartości 0, 1 lub 2, (v) w efekcie zastosowania terapii T uzyskano dwie klasy wyleczenia, czyli $\Omega = \{Y_1, Y_2\}$, co oznacza, że każdy z pacjentów po przeprowadzonej terapii może być scharakteryzowany za pomocą dodatkowego atrybutu r przyjmującego wartości 0 lub 1, które można interpretować: 0 — brak poprawy, 1 — wyleczenie, (vi) wartości atrybutów dla poszczególnych pacjentów przedstawia tabela I.

¹⁶ Przybliżenie idei RST będzie miało charakter podstawowy. Niemniej zakłada się znajomość takich pojęć jak: relacja równoważnościowa, klasa abstrakcji, klasyfikacja, rodzina zbiorów i suma rodziny zbiorów.

¹⁷ Przykładowo dla $f_i(x_i)$, takiego, że $0,3 < f_i(x_i) < 7,8$, przyjmujemy, że $f_i(x_i) = 1$. Dla $f_k(x_k)$, $7,8 \leq f_k$, przyjmujemy, że $f_k = 2$.

Tabela I

	$f_1(a_i)$	$f_2(a_i)$	$f_3(a_i)$	$r(a_i)$
a_1	0	1	0	0
a_2	1	0	2	1
a_3	0	0	0	0
a_4	1	0	2	0
a_5	0	1	0	0
a_6	1	1	1	1
a_7	0	1	1	1
a_8	0	0	1	0

Każdy wiersz w tabeli dostarcza pewnych informacji o danym pacjencie.¹⁸ Przyjmijmy, że $Q = \{f_1, f_2, f_3, r\}$. Zauważmy, że dla dowolnego $A \subseteq Q$ można określić relację nierozróżnialności w zbiorze U . Przykładowo pacjenci a_1, a_5 i a_7 są nierozróżnialni ze względu na zbiór atrybutów $A = \{f_1, f_2\}$. Relacja nierozróżnialności, będąc relacją równoważnościową, wyznacza klasy abstrakcji w zbiorze U .

Gdyby atrybuty f_1, f_2, f_3 stanowiły wszystkie przyczynowo istotne czynniki, to klasy abstrakcji wyznaczone w zbiorze U ze względu na te atrybuty można by jednoznacznie przyporządkować klasom abstrakcji wyznaczonym w zbiorze U ze względu na atrybut r . Wartości atrybutów f_1, f_2, f_3 określałyby wówczas w sposób jednoznaczny wskazania lub przeciwwskazania do zastosowania terapii. W rozważanym przykładzie nie można jednak dokonać takiego przyporządkowania wartości atrybutów do klas wyleczenia. Pacjenci a_2 i a_4 zaburzają spójność uzyskanych wyników. Przy tych samych wartościach atrybutów f_1, f_2, f_3 zastosowanie terapii T doprowadziło u tych pacjentów do uzyskania odmiennych efektów. Jak łatwo się domyślić, problem oceny wagotomii wysoce wybiórczej związany był właśnie z występowaniem tego rodzaju niespójności. Niespójność uzyskanych wyników można scharakteryzować w RST wprowadzając pojęcia: *dolnego przybliżenia zbioru*, *górnego przybliżenia zbioru* i *dokładności przybliżenia*. Punktem wyjścia dla sformułowania tych pojęć będą następujące spostrzeżenia:

- (1) relacja nierozróżnialności ze względu na atrybuty f_1, f_2, f_3 wyznacza następującą klasyfikację Ψ elementów zbioru U : $X_1 = \{a_1, a_5\}$, $X_2 = \{a_2, a_4\}$, $X_3 = \{a_3\}$, $X_4 = \{a_6\}$, $X_5 = \{a_7\}$, $X_6 = \{a_8\}$,

¹⁸ Zauważmy, że rozważany przykład reprezentuje „fikcyjny” przykład Worralla. Większość pacjentów nie odnosi bowiem żadnych korzyści z zastosowanej terapii. Przeprowadzenie badania RCT dokonałoby zatem wyłącznie uśrednienia uzyskanych wyników, wykazując, że terapia nie przynosi oczekiwanych korzyści.

- (2) relacja nierozróżnialności z uwagi na atrybut r wyznacza dwie klasy wyleczenia pacjentów ze zbioru U : $Y_1 = \{a_1, a_3, a_4, a_5, a_8\}$, $Y_2 = \{a_2, a_6, a_7\}$.

Przyjmijmy, że $Y_i \subseteq U$, oraz $K = \{f_1, f_2, f_3\}$. Możemy wówczas powiedzieć, że K -dolnym przybliżeniem zbioru Y_i jest suma tych wszystkich elementów rodziny \mathcal{Y} , które zawierają się w zbiorze Y_i . Zatem dla rozważanego przykładu:

$$\underline{K}Y_1 = \{a_1, a_5\} \cup \{a_3\} \cup \{a_8\}$$

ponieważ:

$$\{a_1, a_5\} \subset \{a_1, a_3, a_4, a_5, a_8\}$$

$$\{a_3\} \subset \{a_1, a_3, a_4, a_5, a_8\}$$

$$\{a_8\} \subset \{a_1, a_3, a_4, a_5, a_8\}$$

Z kolei

$$\underline{K}Y_2 = \{a_6\} \cup \{a_7\}$$

ponieważ:

$$\{a_6\} \subset \{a_2, a_6, a_7\}$$

$$\{a_7\} \subset \{a_2, a_6, a_7\}$$

Dolne przybliżenie zbioru Y_i ze względu na zbiór atrybutów K jest zatem zbiorem tych wszystkich obiektów, które „z pewnością” należą do danej klasy wyleczenia na podstawie opisu dokonanego za pomocą atrybutów ze zbioru K . Zauważmy, że dla pacjentów należących do dolnego przybliżenia zbioru można ustalić algorytm decyzyjny DAD określający warunki podjęcia lub zaniechania terapii.

DAD :

- (reguła 1) Dla dowolnego pacjenta x_i , jeżeli $f_1(x_i) = 1$ lub 0 oraz $f_2(x_i) = 1$ i $f_3(x_i) = 1$, to zastosowanie terapii T powoduje efekt R .
- (reguła 2) Dla dowolnego pacjenta x_i , jeżeli $f_1(x_i) = 0$ oraz $f_2(x_i) = 0$ oraz $f_3(x_i) = 0$ lub $f_1(x_i) = 0$ oraz $f_2(x_i) = 1$ i $f_3(x_i) = 0$ lub $f_1(x_i) = 0$ oraz $f_2(x_i) = 0$ i $f_3(x_i) = 1$, to zastosowanie terapii T powoduje efekt $\neg R$.

Zauważmy, że algorytm decyzyjny DAD nie obejmuje wszystkich pacjentów z tabeli I. Pacjentów a_2 i a_4 nie można bowiem jednoznacznie przyporządkować do dolnego przybliżenia żadnej klasy wyleczenia na podstawie atrybutów ze zbioru K .

W konsekwencji nie można również określić algorytmu *DAD* dla tych przyszłych pacjentów x_i , których wartości atrybutów będą identyczne z wartościami atrybutów pacjentów a_2 i a_4 . Zbiór $\{a_2, a_4\}$ jest obszarem niepewności klasyfikacji.

Wobec przedstawionych do tej pory ustaleń można sformułować następujące zastrzeżenie. Otóż algorytm decyzyjny *DAD* może być również określony na podstawie analizy podgrup po dokonaniu badania RCT. Przeprowadzanie analizy podgrup budzi wprawdzie wątpliwości metodologów [Jaeschke, Cook, Guyatt, 1998b, s. 116], [Yusuf, Wittes, i współ., 1991], ale badanie RCT może być przecież od razu zaplanowane w taki sposób, aby uwzględnić wyłącznie wybrane zakresy wartości pewnych atrybutów np. wiek [Worrall, 2009, s. 296]. Dlaczego zatem mielibyśmy uznać, że analiza danych dokonana za pomocą teorii zbiorów przybliżonych stanowi interesującą alternatywę dla badań RCT? Otóż w przeciwieństwie do badań klinicznych teoria ta pozwala „wychwycić” związki przyczynowo-skutkowe w systemach danych zawierających dużą ilość wyników niespójnych. Realizacja tego celu jest możliwa dzięki temu, że RST pozwala przedstawić zależności pomiędzy wartościami atrybutów „przed terapią” a efektami terapii w kategoriach ilościowych. Pragnąc to uczynić, należy wprowadzić pojęcie *górnego przybliżenia zbioru*.

K -górnym przybliżeniem zbioru Y_i jest suma tych wszystkich elementów rodziny \mathcal{P} , które mają część wspólną z Y_i . Zatem dla rozważanego przykładu:

$$\overline{KY}_1 = \{a_1, a_5\} \cup \{a_3\} \cup \{a_8\} \cup \{a_2, a_4\}$$

$$\overline{KY}_2 = \{a_2, a_4\} \cup \{a_6\} \cup \{a_7\}$$

Korzystając z pojęć dolnego i górnego przybliżenia zbioru, można dokonać ilościowej charakterystyki dokładności przybliżenia zbioru Y_i przez zbiór atrybutów K .

$$\alpha_K(Y) = \frac{\text{card}(KY_i)}{\text{card}(\overline{KY}_i)},$$

gdzie $\text{card}(\dots)$ to liczność odpowiednio dolnego lub górnego przybliżenia zbioru Y_i .

W rozważanym przykładzie można ustalić następujące dokładności przybliżenia dla poszczególnych zbiorów.

$$\alpha_K(Y_1) = 4/6 = 0,66$$

$$\alpha_K(Y_2) = 2/4 = 0,5$$

Widoczne jest, że atrybuty ze zbioru K z większą dokładnością charakteryzują pacjentów, którym terapia T nie przyniesie korzyści, niż pacjentów, dla których terapia ta byłaby korzystna.

8.

Ilościowa charakterystyka dokładności przybliżenia zbioru Y przez zbiór atrybutów K może być w RST uogólniona dla klasyfikacji. Posługujemy się wówczas pojęciem *jakości klasyfikacji*. Ogólnie można powiedzieć, że jakością klasyfikacji Ω jest stopień przybliżenia klasyfikacji wyznaczonej przez wartości atrybutów ze zbioru K do klasyfikacji wyznaczonej przez wartości atrybutu r . Jest to zatem stosunek wszystkich „poprawnie” zaklasyfikowanych obiektów do ilości wszystkich obiektów z U . „Poprawnie” zaklasyfikowane obiekty to te, które są elementami dolnych przybliżeń zbiorów. Pojęcie jakości klasyfikacji można zatem sprecyzować następująco.

Niech χ będzie rodziną zbiorów, której elementami są $\underline{K}Y_1$ oraz $\underline{K}Y_2$ natomiast $\Sigma\chi$ oznacza sumę rodziny χ . Wówczas jakość klasyfikacji Ω ze względu na atrybuty ze zbioru K określimy:

$$\kappa(\Omega) = \frac{\text{card}(\Sigma\chi)}{\text{card}(U)} = 6/8 = 0,75$$

Ilościowa charakterystyka jakości klasyfikacji stanowi punkt wyjścia dla ustalenia zależności przyczynowych pomiędzy poszczególnymi atrybutami ze zbioru K a efektami terapii. Idea jest następująca. Jeżeli dany atrybut jest przyczynowo istotny dla efektów terapii, to usunięcie tego atrybutu ze zbioru K spowoduje, że obniży się dokładność przybliżeń dla poszczególnych klas wyleczenia, co z kolei doprowadzi do obniżenia jakości całej klasyfikacji. Przyjmijmy, że w odniesieniu do rozważanego przykładu usunięty zostanie atrybut f_2 . W konsekwencji uzyskujemy zbiór atrybutów $K' = \{f_1, f_3\}$. Dla zbioru atrybutów K' można ustalić dolne przybliżenia klas wyleczenia Y_1 i Y_2 oraz jakość klasyfikacji Ω ze względu na K' .

$$\underline{K}'Y_1 = \{a_1, a_5, a_3\}$$

$$\underline{K}'Y_2 = \{a_6\}$$

Współczynnik jakości klasyfikacji po usunięciu atrybutu f_2 wynosi:

$$\kappa(\Psi) = 4/8 = 0,5$$

Nastąpiło zatem obniżenie jakości klasyfikacji z wartości 0,75 do wartości 0,5. Na tej podstawie wnosimy, że atrybut f_2 był przyczynowo istotny dla efektów terapii.

Oczywiście może się również zdarzyć, że usunięcie danego atrybutu ze zbioru K nie doprowadzi do obniżenia jakości klasyfikacji. Wówczas wnosimy, że atrybut ten nie jest przyczynowo istotny dla efektów terapii.¹⁹

¹⁹ W teorii zbiorów przybliżonych sprecyzowano pojęcia umożliwiające analizę tego rodzaju sytuacji. Z uwagi na wstępny charakter artykułu pojęcia te nie zostaną wprowadzone.

9.

Zauważmy, że zastosowanie RST w analizie danych medycznych pozwala sformułować prosty algorytm oceny danych. Załóżmy, że dysponujemy skończonym zbiorem atrybutów $K = \{f_1, \dots, f_i\}$. Usunięcie różnych atrybutów ze zbioru K może prowadzić do obniżenia jakości klasyfikacji o różne wartości. Jeżeli usunięcie atrybutu f_k prowadzi do większego obniżenia jakości klasyfikacji, niż usunięcie atrybutu f_n , to uznajemy, że atrybut f_k jest przyczynowo bardziej istotny dla efektów terapii niż atrybut f_n . Przy dużej ilości atrybutów i dużej liczbie przebadanych pacjentów zastosowanie metody RST w analizie danych wymaga wykorzystania odpowiednich programów komputerowych i dysponowania odpowiednią mocą obliczeniową.²⁰

W ocenie danych empirycznych dotyczących efektów terapii wagotomii wysoce wybiórczej u 122 pacjentów brano pod uwagę 11 atrybutów. Ustalona klasyfikacja zawierała 4 klasy wyleczenia. Stwierdzono, że usunięcie dwóch atrybutów (płeć i współczynnik BAO) nie powoduje obniżenia jakości klasyfikacji [Słowiński, 1990, s. 31]. Uznano zatem, że nie są one przyczynowo istotne dla efektów terapii. Ze zbioru zawierającego pozostałe 9 atrybutów usuwano kolejne atrybuty i sprawdzano, o jaką wartość obniża się jakość klasyfikacji. Ostatecznie wyodrębniono zbiór 5 atrybutów, których usuwanie prowadziło do największego obniżenia jakości klasyfikacji. Różne wartości tych atrybutów posiadały decydujący wpływ na efekty terapii. Stanowiły one o wskazaniach i przeciwwskazaniach do wykonywania operacji. Przeprowadzona analiza pozwoliła zatem ustalić algorytm decyzyjny o charakterze deterministycznym. Precyzował on, jakie wartości atrybutów przesądzają o tym, do której klasy wyleczenia będzie należał dany pacjent po zastosowaniu terapii T.

10.

Teoria zbiorów przybliżonych jest jedną z wielu metod stosowanych współcześnie w analizie danych empirycznych w medycynie. Stanowi ona niewątpliwie interesującą alternatywę wobec statystycznych metod pozyskiwania danych z badań klinicznych, w tym również badań RCT. Pozwalając ustalić związki przyczynowe pomiędzy efektami terapii a wartościami czynników zakłócających, RST staje się szczególnie ważnym narzędziem oceny danych w przypadku zaobserwowania niezgodnych efektów terapii.

Dla filozofa nauki zastosowanie teorii zbiorów przybliżonych w ocenie danych medycznych może być szczególnie interesującym przedmiotem analiz epistemologicznych. Wskazać można na kilka ważnych problemów. Po pierwsze, zastosowanie

²⁰ Dla potrzeb analizy danych medycznych w oparciu o teorię zbiorów przybliżonych opracowane zostały programy komputerowe ROSE (Rough Set Data Explorer) i ProFIT (Rough Processing of Fuzzy Information Tables). Umożliwiają one analizę danych medycznych i tworzenie algorytmów decyzyjnych [Słowiński, 2003, s. 411].

teorii zbiorów przybliżonych dla oceny danych empirycznych wymaga założenia, że dysponujemy kryterium wstępnej selekcji atrybutów, które są dołączane do zbioru K . Mają to być mianowicie te atrybuty, które rozpoznajemy jako potencjalnie przyczynowo istotne dla efektów terapii. Sama teoria zbiorów przybliżonych takiej metody selekcji atrybutów nie dostarcza. Postawić można wówczas pytanie o wiarygodność kryterium wstępnego przydziału atrybutów do zbioru K . Drugi problem dotyczy usuwania ze zbioru K atrybutów, które w najmniejszym stopniu obniżają jakość klasyfikacji. Zauważmy, że nie są to atrybuty nieistotne przyczynowo, lecz takie, które mają najmniejszy wpływ na efekt terapii. Jeżeli jednak atrybuty te mają jakiś wpływ, to usunięcie ich ze zbioru K ma charakter arbitralny. Stosowany wówczas w analizach medycznych termin *zadowalającej jakości klasyfikacji* [Słowiński, 1990], [Słowiński, 2003] jest terminem epistemologicznym i wymaga sprecyzowania właśnie na terenie epistemologii.

Wskazane zagadnienia nie wyczerpują wszystkich problemów epistemologicznych związanych z zastosowaniem teorii zbiorów przybliżonych w ocenie danych medycznych. Problemy te konstytuują nowy obszar analiz, które mogą być podjęte we współczesnej filozofii nauki.

BIBLIOGRAFIA

- Abel U., Koch A., (1999), „The Role of Randomization in Clinical Studies: Myths and Beliefs”, *Journal of Clinical Epidemiology*, vol. 52(6), s. 487-497.
- Brzeziński Z., Szamotulska K., (1997), *Epidemiologia kliniczna*, Wydawnictwo Lekarskie PZWL, Warszawa.
- Cartwright N., (1989), *Nature's Capacities and Their Measurement*, Clarendon Press, Oxford.
- Cartwright N., (2010), „What are randomized controlled trials good for?”, *Philosophical Studies* 147, s. 59-70.
- Cartwright N., Munro E., (2010), „The limitations of randomized controlled trials in predicting effectiveness”, *Journal of Evaluation in Clinical Practice*, 16, s. 260-266.
- Guyatt G., Jaeschke R., Cook D. i współ. (2000), „Evidence-based medicine (EBM) współczesną sztuką lekarską. Cele Polskiego Instytutu Evidence Based Medicine”, *Medycyna Praktyczna*.
- Gajewski P., Jaeschke R., i współ., (2003), „Evidence based medicine (EBM) współczesną sztuką lekarską. Cele Polskiego Instytutu Evidence Based Medicine”, *Medycyna Praktyczna*, 3(145), s. 31-34.
- Hacking I., (1988), „Telepathy: Origins of Randomization In Experimental Design”, *Isis*, 79, s. 427-51.
- Jaeschke R., Cook D., Guyatt G., (1998a), „Evidence based medicine (EBM), czyli praktyka medyczna oparta na wiarygodnych i aktualnych publikacjach (POWAP)”, *Medycyna Praktyczna*, 2(84), s. 149-155.
- Jaeschke R., Cook D., Guyatt G., (1998b), „Evidence based medicine (EBM), czyli praktyka medyczna oparta na wiarygodnych i aktualnych publikacjach (POWAP)”, *Medycyna Praktyczna*, 5(87), s. 113-121.
- Kuipers T., Wiśniewski A., (1994), „An Erotetic Approach to Explanation by Specification”, *Erkenntnis* 40, s. 377-402.

- Papineau D., (1994), „The Virtues of Randomization”, *The British Journal for the Philosophy of Science*, vol. 45(2), s. 437-450.
- Pawlak Z., Słowiński K i współ. (1986), Rough classification of patients after highly selective vagotomy for duodenal ulcer”, *International Journal of Man-Machine Studies*, 24, s. 413-433.
- Roush S., (2009), „Randomized controlled trials and the flow of information: comment on Cartwright”, *Philosophical Studies* 143, s. 137-145.
- Słowiński K., (1990) „Wykorzystanie teorii zbiorów przybliżonych do analizy leczenia wrzodu dwunastnicy wysoce wybiórczą wagotomią i ostrego zapalenia trzustki płukaniem otrzewnej”, Poznań.
- Słowiński K., (2003) „Zbiory przybliżone”, [w:] *Kompendium informatyki medycznej*, (ed.) R. Zajdel, E. Kaćki i współ., Bielsko-Biała, α -medica Press, s. 403-421.
- Straus S., Richardson W., i współ. (2005), *Evidence-Based Medicine: How to Practice and Teach*, Elsevier Churchill Livingstone.
- Urbach P., (1985), „Randomization and the Design of Experiments”, *Philosophy of Science*, vol. 52(2), s. 256-273.
- Worrall J., (2002), „What Evidence in Evidence-Based Medicine?”, *Philosophy of Science*, vol. 69(S3), s. 316-330.
- Worrall J., (2007), „Why There’s No Cause to Randomize”, *British Journal for the Philosophy of Science*, 58, s. 451-488.
- Worrall J., (2009), „Do we need some large, simple randomized trials in medicine”, [w:] *EPSA Philosophical Issues in the Science*, Suarez M., Dorato M., Redei M., (ed.), Heidelberg, London, New York, s. 289-303.
- Wulff H., Götzsche P., (2005), *Racjonalna diagnoza i leczenie. Wprowadzenie do medycyny wiarygodnej, czyli Evidence-Based Medicine*, Łódź: Aktis.
- Yusuf S., Wittes J., i współ. (1991), „Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials”, *Journal of American Medical Association*; 266, s. 93-98.