

TOMASZ RZEPIŃSKI

STRUKTURY WYJAŚNIANIA STATYSTYCZNEGO  
WARUNEK ISTOTNOŚCI STATYSTYCZNEJ W ANALIZIE PODGRUP  
RANDOMIZOWANYCH BADAŃ KLINICZNYCH

1. WSTĘP

Przedmiotem analiz prowadzonych w refleksji nad nauką są dwa główne obszary. Pierwszy obejmuje wytwory procesu poznania naukowego: teorie, modele teoretyczne, prawa i twierdzenia identyfikujące związki przyczynowe, statystyczne, funkcjonalne itp. Drugi obszar obejmuje procedury badawcze stosowane w nauce: wyjaśnianie, sprawdzanie hipotez, falsyfikację, konfirmację, formułowanie hipotez badawczych i wiele innych. Elementy obu tych obszarów określane są mianem struktur metodologicznych. Struktury te, jak pisze Zygmunt Hajduk, stanowią podstawowe jednostki analizy metodologicznej (HAJDUK 2016, 7).

Zagadnienie wyjaśniania stanowiło przez długi czas jeden z najistotniejszych problemów podejmowanych w filozoficznej refleksji nad nauką. Możliwość sformułowania wyjaśnień świadczy bowiem o zrozumieniu procesów, w których zachodzą badane zjawiska. Ten aspekt podkreślany jest m.in. przez Z. Hajduka, piszącego, że wysiłek badawczy podejmowany w nauce jest „[...] skierowany na wyjaśnianie; dążymy do zrozumienia. Wynikiem poznania naukowego ma być rozumienie” (HAJDUK 2016, 109). Problem polega jednak na tym, że procedura wyjaśniania nie jest jednolita. Formułowane wyjaśnienia przyjmują różną postać w różnych dyscyplinach nauki. Wyróżnienie w zakresie analiz metodologicznych odmiennych typów wyjaś-

---

Dr hab. TOMASZ RZEPIŃSKI, prof. UAM — Zakład Logiki i Metodologii Nauk, Uniwersytet im. A. Mickiewicza w Poznaniu oraz Katedra Biologii i Ochrony Środowiska, Uniwersytet Medyczny im. K. Marcinkowskiego w Poznaniu; adres do korespondencji: ul. Szamarzewskiego 89c, 60-568 Poznań; e-mail: rzeto@amu.edu.pl

niania wydaje się wskazywać, że zrozumienie stanowiące cel poznawczy uczonych może być realizowane w różny sposób. Problemem nie jest jednak wyłącznie to, że w różnych naukach dominują odmienne typy wyjaśniania. Problem polega na tym, że w przypadku niektórych typów wyjaśniania nie ma zgody metodologów co do tego, jakie cele poznawcze pozwalają one w ogóle realizować. Często nie ma również zgody co do tego, kiedy dany typ wyjaśniania jest poprawny.

Wydaje się, że najwięcej zastrzeżeń metodologów, zarówno jeżeli chodzi o warunki poprawności, jak i realizowane cele poznawcze, budzi wyjaśnianie statystyczne. Ten typ wyjaśniania podejmowany jest w wielu naukach, zarówno społecznych, jak i przyrodniczych. W artykule omówiony zostanie spór dotyczący zadań poznawczych realizowanych w procesie wyjaśniania statystycznego podejmowanego w badaniach biomedycznych. Punktem wyjścia dla prowadzonych rozważań będzie przypomnienie modelu wyjaśniania indukcyjno-statystycznego opracowanego przez Carla G. Hempla. Krytyka formułowana wobec proponowanego w nim ujęcia wyjaśniania stanowiła, jak wiadomo, podstawę dla rozwinięcia modelu istotności statystycznej zaproponowanego przez Wesleya C. Salmona. Zasadniczym celem artykułu będzie pokazanie, że warunek istotności statystycznej zawodzi i jest nie do zaakceptowania jako warunek określający poprawność wyjaśnień statystycznych. W szczególności będę argumentował na rzecz tezy, że w przeciwieństwie do innych typów wyjaśnień ocena poprawności wyjaśniania statystycznego wymaga analizy procedur badawczych stanowiących podstawę formułowania twierdzeń tworzących eksplanans. Dla potrzeb realizacji wskazanego celu zaprezentuję szczególny typ badań biomedycznych, jakim jest analiza podgrup randomizowanych badań klinicznych.

## 2. DWA KONTEKSTY WYJAŚNIANIA

Procedura wyjaśniania ma na celu sformułowanie odpowiedzi na pytanie o postaci „Dlaczego Z?”, w którym Z, określane mianem *eksplanandum*, jest twierdzeniem opisującym dane zjawisko. *Eksplanans* jest natomiast koniunkcją zdań dostarczających wyjaśnienia zjawiska opisanego w zdaniu Z.

Charakteryzując procedurę wyjaśniania, należy wyraźnie odróżnić dwie sytuacje, w których jest ona przeprowadzana. W pierwszej wyjaśnienie danego zjawiska wymaga tylko przywołania odpowiednich teorii, praw lub hipotez, które zostały już wcześniej sformułowane w nauce i które posiadają nie-

zależną ewidencję empiryczną. Z. Hajduk ujmuje tę kwestię w następujący sposób: „Wyjaśnienie jest też nieraz charakteryzowane jako wiązanie tego, co nieznanne, z tym, co znane, albo też przez ukazywanie, jak (w jaki sposób) to, co nieznanne (niezrozumiałe), można sprowadzić do tego, co już znane bądź zrozumiałe. Tłumaczenia czynią zdarzenia oczekiwanymi, a nie zaskakującymi” (HAJDUK 2016, 30).

W drugiej sytuacji wyjaśnianie podejmowane jest wówczas, gdy nie dysponujemy prawami lub hipotezami, w oparciu o które można wyjaśnienie sformułować. W takim wypadku procedura wyjaśniania ma charakter wiedzotwórczy, przyczyniając się do generowania nowych hipotez badawczych w nauce. Zwykle ten typ wyjaśniania ma miejsce w sytuacjach, w których zaobserwowano zjawiska zaskakujące, abnormalne<sup>1</sup> (RZEPIŃSKI 2011).

Przedmiotem analiz metodologicznych może być zarówno struktura eksplanansu, stanowiącego odpowiedź na sformułowane pytanie o wyjaśnienie, jak i proces wnioskowania prowadzący do sformułowania wyjaśnień. Pierwszy z tych aspektów wyjaśniania sprecyzowany został w ramach dedukcyjno-nomologicznego modelu wyjaśniania, zaproponowanego w latach 40. XX wieku przez Carla G. Hempela i Paula Oppenheima (HEMPEL i OPPENHEIM 1948). W koncepcji tej uściśła się, na czym polega poprawność wyjaśniania naukowego poprzez określenie warunków, które powinien spełniać eksplanans. Pierwszy z nich ustala, że eksplanans powinien zawierać prawa ogólne oraz zdania stwierdzające spełnienie warunków początkowych<sup>2</sup>. Drugi określa, że eksplanandum powinno wynikać logicznie z eksplanansa. Trzeci natomiast stwierdza, że eksplanans powinien posiadać niezależną od eksplanandum ewidencję empiryczną, co pozwala uniknąć błędu wyjaśniania *ad hoc*.

Zauważmy, że w modelu Hempela-Oppenheima ustalenie poprawności wyjaśniania nie wymaga analizy poprawności procedur badawczych, które pozwoliły na sformułowanie twierdzeń tworzących eksplanans. W artykule będę starał się pokazać, że wyjaśnianie statystyczne podejmowane w naukach biologiczno-medycznych nie spełnia tego warunku. Analiza procedur badawczych stanowiących podstawę formułowania twierdzeń statystycznych przywoływanych w eksplanansie jest nieodzownym elementem oceny po-

<sup>1</sup> Odkrywczy kontekst procedury wyjaśniania coraz częściej staje się przedmiotem analiz filozofów nauki. Z jednej strony zainteresowanie to jest inicjowane rozwojem logik erotetycznych, które dostarczają narzędzi pozwalających na dokonywanie tego typu rozważań. Z drugiej natomiast zainteresowanie to jest konsekwencją zwrotu, jaki obecnie dokonuje się w metodologii w stronę badań typu *case studies*. Przeprowadzane analizy tego typu wyraźnie wskazują na wiedzotwórczy mechanizm wyjaśniania w rozwoju nauki.

<sup>2</sup> Ten koniunkcyjny warunek często jest prezentowany pod postacią dwóch osobnych twierdzeń.

prawności formułowanych wyjaśnień statystycznych. W skrajnych sytuacjach twierdzenia identyfikujące te same zależności statystyczne mogą być rozpoznawane jako twierdzenia pełniące odmienne funkcje poznawcze: bądź jako eksplanans w wyjaśnianiu zależności statystycznych, bądź jako twierdzenia generujące nowe hipotezy badawcze. W tym drugim wypadku pełnią one funkcję wiedzotwórczą. Podkreślić należy, że rozpoznanie statusu poznawczego twierdzenia statystycznego nie zależy od jego struktury i treści, lecz od procedury badawczej, która doprowadziła do jego sformułowania. Tego aspektu wyjaśniania statystycznego nie ujmuje zarówno model indukcyjno-statystyczny C.G. Hempla, jak i model istotności statystycznej W. Salmona. Wydaje się tymczasem, że kwestia ta jest istotna, ponieważ wskazuje na ważną cechę tego typu wyjaśnień. Są one mianowicie zależne od innych procedur badawczych. Konsekwencją przyjęcia takiego stanowiska jest uznanie, że nie można, wbrew zamiarom C.G. Hempla i W. Salmona, określić warunków poprawności wyjaśnień statystycznych wyłącznie na podstawie struktury tych wyjaśnień.

### 3. STRUKTURA TWIERDZEŃ STATYSTYCZNYCH

Analiza struktury twierdzeń statystycznych przeprowadzana w metodologii ma na celu rozstrzygnięcie trzech kwestii. Pierwszą jest ustalenie budowy tych twierdzeń, drugą — rozróżnienie między prawami a generalizacjami statystycznymi i ocena doniosłości obu tych rodzajów twierdzeń w naukach przyrodniczych, trzecią natomiast sprecyzowanie celów poznawczych realizowanych za pomocą tych twierdzeń w nauce.

Najogólniej twierdzenia występujące w eksplanansie wyjaśnień statystycznych to twierdzenia, które przyporządkowują wystąpieniu pewnych rodzajów zdarzeń pewne wartości prawdopodobieństwa. Są to zatem twierdzenia w rodzaju:

- (i) Prawdopodobieństwo wyrzucenia reszki przy jednokrotnym rzucie monetą wynosi 0,5.
- (ii) Prawdopodobieństwo wystąpienia zdarzenia niekorzystnego w populacji P osób leczonych terapią I wynosi  $p$ .
- (iii) Prawdopodobieństwo wystąpienia choroby D u osoby O przy pozytywnym wyniku testu T wynosi  $p$ .

- (iv) Prawdopodobieństwo wystąpienia osobników homozygotycznych z cechą recesywną w fenotypie wśród potomstwa pochodzącego od organizmów heterozygotycznych wynosi 0,25.
- (v) Przy 25-krotnym wyrzuceniu dwóch kostek do gry prawdopodobieństwo jednoczesnego wyrzucenia dwóch szóstek jest większe niż 0,5.

Zauważmy, że twierdzenia te można podzielić na dwie grupy. Pierwsza obejmuje takie twierdzenia, które określają wartość prawdopodobieństwa wystąpienia danego zdarzenia niezależnie od danych empirycznych. Są to: (i), (iv) oraz (v). W przypadku każdego z tych twierdzeń można ustalić wartość prawdopodobieństwa, o ile w odpowiedni sposób rozumiemy pojęcie: moneta, homozygota, heterozygota, kostka. Warunkiem ustalenia prawdopodobieństwa wystąpienia danego zdarzenia jest w tym przypadku odpowiedni sposób rozumienia samego pojęcia prawdopodobieństwa jako stosunku względnej częstości zdarzeń sprzyjających do wszystkich możliwych zdarzeń (AJDUKIEWICZ 1975). O takim pojęciu prawdopodobieństwa mówimy, że ma charakter *a priori*. Ustalenie jego wartości nie wymaga odwoływania się do doświadczenia, a co najwyżej przyjęcia określonych założeń dotyczących badanych przedmiotów lub dokonania pewnych matematycznych przekształceń<sup>3</sup>.

Twierdzenia, w których występuje pojęcie prawdopodobieństwa *a priori*, określane są w literaturze przedmiotu mianem *statystycznych twierdzeń czasoprzestrzennie otwartych* (KMITA 1975, 168]. Uznaje się, że pełnią one funkcję praw nauki<sup>4</sup>. Jak pisze W. Salmon: „Prawa prawdopodobieństwa nie są prawami empirycznymi; są prawami matematyki, które nie posiadają zawartości faktualnej” (SALMON 2006, 51). Ich zastosowanie do ustaleń nauk empirycznych, jak w przypadku twierdzenia (iv), stanowi wyłącznie efekt zastosowania generalizacji statystycznych w odniesieniu do określonego obszaru badań.

Drugi rodzaj twierdzeń statystycznych to twierdzenia, które ustalają wartość prawdopodobieństwa wystąpienia określonych rodzajów zdarzeń na podstawie danych doświadczenia. W twierdzeniach tych posługujemy się zatem pojęciem prawdopodobieństwa *a posteriori*. Wśród podanych przykładów są

---

<sup>3</sup> Na przykład w twierdzeniu (v) zakładamy, że każdy z osobników heterozygotycznych ma kombinację genów Aa lub aA oraz że w efekcie ich skrzyżowania uzyskujemy cztery możliwe kombinacje genów, z których tylko jedna jest zdarzeniem sprzyjającym. Z kolei w twierdzeniu (vi) posługujemy się dwumianem Newtona i próbą Bernoulliego (GILLIES 2010, 5).

<sup>4</sup> Czyli statystycznych twierdzeń ściśle ogólnych. Ścisła ogólność oznacza czasoprzestrzenną otwartość tych twierdzeń.

to twierdzenia (ii) oraz (iii)<sup>5</sup>. W literaturze przedmiotu tego rodzaju twierdzenia określane są mianem statystycznych generalizacji historycznych, ponieważ odnoszą się do czasoprzestrzennie zamkniętych zbiorów przedmiotów (KMITA 1975, 169). Nie są to zatem prawa nauki.

Próba odróżnienia praw nauki od generalizacji historycznych za pomocą pojęcia czasoprzestrzennej otwartości wyraża niewątpliwie bardzo ważną, metodologiczną, intuicję. Inna jest przecież doniosłość poznawcza twierdzenia ogólnego, ustalającego, że wszystkie ciała podlegają sile grawitacji, a inna twierdzenia, że wszystkie renesansowe dzieła sztuki reprezentują typ linearny (KMITA 1975, 166). Niestatystyczne twierdzenia ściśle ogólne są niewątpliwie bardziej doniosłe poznawczo niż generalizacje historyczne. Zauważmy jednak, że ustalenie to budzi już wątpliwości w odniesieniu do twierdzeń statystycznych.

Statystyczne twierdzenia ściśle ogólne są czasoprzestrzennie otwarte z uwagi na przyjęty sposób rozumienia prawdopodobieństwa jako prawdopodobieństwa *a priori* w interpretacji statystycznej. Problem polega jednak na tym, że interpretacja statystyczna stosuje się wyłącznie do skończonych zbiorów zdarzeń, wyczerpujących rozważane uniwersum zdarzeń. Jeżeli rzucamy monetą, kością, wyciągamy kartę z talii kart, to możemy ustalić wartość prawdopodobieństwa w interpretacji statystycznej, ponieważ zbiory rozważanych zdarzeń są skończone (odpowiednio obejmują 2 zdarzenia, 6 zdarzeń lub 52). Ponadto są to zbiory wyczerpujące uniwersum rozważanych zdarzeń. Nie ma takich dodatkowych zdarzeń, które byłyby jeszcze przez nas rozważane<sup>6</sup>. O takich zbiorach zdarzeń mówimy, że mają charakter ruletkowy.

Prawa ustalające zależności statystyczne nie uwzględniają natomiast sytuacji, w których zbiory zdarzeń o mocy dążącej do nieskończoności lub też sytuacji, w których uczony formułuje twierdzenie statystyczne wyłącznie na podstawie znajomości pewnej skończonej liczby przypadków niewyczerpujących rozważanego uniwersum zdarzeń, np. (ii) lub (iii). Problem polega na tym, że w naukach społecznych i przyrodniczych najczęściej mamy do czynienia właśnie z tą drugą sytuacją. Na podstawie pewnej ilości znanych zdarzeń określa się wartość prawdopodobieństwa. Tak właśnie jest w przy-

<sup>5</sup> Wartość prawdopodobieństwa w twierdzeniu (ii) określana jest za pomocą pojęć ryzyka i względnej redukcji ryzyka na podstawie danych uzyskanych z badań biomedycznych (RZEPIŃSKI 2016). Z kolei wartość prawdopodobieństwa w twierdzeniu (iii) określana jest jako wartość predykcji lub z wykorzystaniem analizy bayesowskiej również na podstawie danych empirycznych (RZEPIŃSKI [w druku]).

<sup>6</sup> Innymi słowy wykluczamy sytuację, w której moneta lub kostka zatrzymają się na swojej krawędzi.

padku twierdzeń (ii) oraz (iii). To właśnie te twierdzenia są doniosłe poznawczo. Stanowią podstawę formułowanych wyjaśnień i przewidywań.

W kontekście dokonanych ustaleń nasuwa się ważne spostrzeżenie. Otóż twierdzenia statystyczne będące generalizacjami statystycznymi są znacznie częściej wykorzystywane w naukach biomedycznych aniżeli statystyczne twierdzenia ściśle ogólne w rodzaju praw Mendla — (iv). Innymi słowy, w naukach tych znacznie częściej wykorzystywane są twierdzenia statystyczne operujące prawdopodobieństwem *a posteriori* aniżeli prawa, w których przyjmuje się statystyczną interpretację prawdopodobieństwa.

Na zakończenie pozostaje jeszcze ustalić budowę twierdzeń statystycznych. W. Salmon wyróżnia dwie możliwości. Pierwsza jest następująca:

*Każdy przedmiot posiadający cechę F posiada skłonność [propensity] do bycia przedmiotem posiadającym cechę G (SALMON 2006, 52).*

Wydaje się, że taka struktura pozwala uwzględnić intuicje wyrażane we wszystkich twierdzeniach, w których akceptuje się statystyczną interpretację prawdopodobieństwa, tj. (i), (iv) oraz (v). Przy tej interpretacji opisujemy bowiem skłonność przedmiotów do udziału w pewnych zdarzeniach niezależnie od danych doświadczenia. Trudność może jednak sprawić ustalenie, co jest przedmiotem kwantyfikacji. O ile bowiem dla twierdzeń (i) oraz (iv) mogą to być przedmioty konkretne (monety, pary kości), o tyle dla (v) musiałyby to być zdarzenia.

Drugi typ wyróżnionych przez W. Salmona twierdzeń, które mogą kandydować do miana twierdzeń statystycznych, ma postać:

*Pewna część przedmiotów posiadających cechę F posiada cechę G (SALMON 2006, 52).*

Formułując tego rodzaju twierdzenia, pragniemy — jak pisze W. Salmon — uwzględnić fakt, że wśród wszystkich przedmiotów typu F są takie, które mają cechę G. Wydaje się zatem, że w tym przypadku struktura tych twierdzeń pozwalałaby ująć intuicję przyjmowaną przy posługiwaniu się pojęciem prawdopodobieństwa *a posteriori* w twierdzeniach opisujących proporcje występowania danej cechy aktualnie, np. twierdzenie (iii) — Pewna część osób z pozytywnym wynikiem testu diagnostycznego jest chora<sup>7</sup>.

Obydwa sformułowania są, w moim przekonaniu, dosyć kontrowersyjne i trudno jest je zaakceptować. Wydaje się, że intuicje, jakie wiążemy z twierdzeniami statystycznymi, w większym stopniu zostały uwzględnione w uję-

<sup>7</sup> Dla potrzeb bardziej precyzyjnej analizy tego przykładu odsyłam do RZEPIŃSKI (w druku).

ciu Jerzego Kmity. Jego zdaniem prawa statystyczne są twierdzeniami o postaci:

*Względna częstość cechy  $F$  wyznaczającej zbiór  $A$  w zbiorze  $L$  wynosi  $p$ , gdzie  $p$  – to wartość prawdopodobieństwa (KMITA 1975, 168).*

Taka struktura ujmuje zarówno twierdzenia, które są statystycznymi generalizacjami historycznymi, jak i twierdzenia będące prawami statystycznymi.

#### 4. CARLA G. HEMPLA KONCEPCJA WYJAŚNIANIA STATYSTYCZNEGO I JEJ KRYTYKA

Analizując strukturę wyjaśnień statystycznych, C.G. Hempel odróżnił dwa główne ich typy. Pierwszym jest wyjaśnianie dedukcyjno-statystyczne. Ma ono miejsce wówczas, gdy przedmiotem wyjaśniania jest prawidłowość statystyczna. Eksplanans zawiera wówczas szersze prawo statystyczne, z którego wynika dedukcyjnie prawidłowość statystyczna stanowiąca eksplanandum. Drugim jest wyjaśnianie indukcyjno-statystyczne. Mamy z nim do czynienia wówczas, gdy eksplanandum jest zdaniem opisującym pojedyncze zdarzenie, natomiast eksplanans zawiera przynajmniej jedno twierdzenie statystyczne wskazujące na wysokie prawdopodobieństwo wystąpienia pewnego typu zdarzeń.

Wyjaśnianie dedukcyjno-statystyczne ilustrowane jest zwykle następującym przykładem eksperymentów z rzutem kośćmi do gry, które przeprowadził Chavalier de Méré w XVII wieku. Badał on dwie sytuacje. W pierwszej, na podstawie obserwacji określał względną częstość wyrzucenia przynajmniej jednej szóstki przy czterokrotnym rzucie kostką. Stwierdził, że prawdopodobieństwo wyrzucenia przynajmniej jednej szóstki w opisanej sytuacji wynosi powyżej 50%. Druga analizowana przezeń eksperymentalnie sytuacja dotyczyła ustalenia prawdopodobieństwa wyrzucenia dwóch szóstek przy jednoczesnym rzucie dwiema kostkami do gry (GILLIES 2010, 93). Okazało się, że aby wartość prawdopodobieństwa była większa niż 0,5, należy dokonać 25 rzutów.

Pytanie o wyjaśnienie sformułowane przez de Méré miało zatem postać: dlaczego, rzucając dwiema kostkami do gry 24 razy, prawdopodobieństwo przynajmniej jednokrotnego wyrzucenia pary szóstek jest mniejsze niż 50%? (SALMON 2006, 51). Wyjaśnienie dedukcyjno-statystyczne wymagałoby, aby w eksplanansie znajdowało się prawo statystyczne, z którego dedukcyjnie



wyprowadzalne byłoby eksplanandum. Za czasów de Méré wyjaśnienie to nie mogło być jednak podane. Prawo takie zostało sformułowane dopiero w XVIII wieku. Była to tzw. próba Bernoulliego, pozwalająca określić prawdopodobieństwo  $k$ -ilości sukcesów w  $n$ -ilości prób.

Omówiony przykład dobrze ilustruje sytuację, w której podejmowanie wyjaśnień statystycznych prawidłowości empirycznych prowadzi do sformułowania praw o większym stopniu ogólności aniżeli twierdzenia opisujące prawidłowości zaobserwowane. W omawianym przykładzie prawo o większym stopniu ogólności zakładało statystyczną interpretację prawdopodobieństwa *a priori*. Rozważmy odmienną sytuację.

Przyjmijmy, że biolog obserwuje populację kilkunastu osobników będących potomstwem organizmów heterozygotycznych. Na podstawie obserwacji stwierdza, że względna częstość występowania w tej populacji homozygot o cesze recesywnej w fenotypie wynosi 0,4. Na pytanie o wyjaśnienie tej wartości biolog nie może odwołać się wyłącznie do praw genetyki, ponieważ one ustalają wartość 0,25. Zaobserwowana prawidłowość empiryczna nie może być zatem wyjaśniona w modelu dedukcyjno-statystycznym na podstawie prawa ustalającego prawdopodobieństwo statystyczne *a priori*. Uczony może odwołać się do czynników natury biologicznej, które wpłynęły na zwiększenie liczby homozygot recesywnych, lub uznać, że badana próbka jest zbyt mała. W tym drugim przypadku uznaje zatem, że zwiększenie liczby obserwowanych osobników pozwoli stwierdzić zależność opisaną w prawach statystycznych formułowanych na gruncie genetyki. Opisana sytuacja została uwzględniona przez Von Misesa w jego częstościowej interpretacji prawdopodobieństwa. Uznał on, że obserwacje ciągów zdarzeń pozwalają stwierdzić proces stabilizowania statystycznych częstości wraz ze zwiększającą się liczbą obserwowanych zdarzeń. Spostrzeżenie to D. Gillies określa mianem prawa stabilizowania statystycznych częstości (*Law of Stability of Statistical Frequencies* — LSSF) (GILLIES 2010, 92). Zdaniem Von Misesa prawo to znajduje potwierdzenie we wszystkich grach, w których możliwe jest określenie prawdopodobieństwa statystycznego (kości, ruletka, karty, loteria) oraz w zdarzeniach analizowanych w naukach biologicznych (GILLIES 2010, 92)<sup>8</sup>. Prawo LSSF stanowi zatem niezbędny element, który musi być przywołany w przypadku wyjaśniania dedukcyjno-statystycznego w sytuacjach, w których analizowany jest zbyt mało liczny zbiór zdarzeń. Innymi słowy, w sytuacjach, w których wartość prawdopodobieństwa oceniona

<sup>8</sup> Proces stabilizowania statystycznych częstości Von Mises określa mianem *podstawowego zjawiska teorii prawdopodobieństwa (primary phenomenon theory of probability)* (GILLIES 2010).

*a posteriori* jest niezgodna z wartością *a priori* ustalaną na podstawie bardziej ogólnych praw statystycznych.

Przejdźmy do omówienia struktury wyjaśnień *indukcyjno-statystycznych* (I-S). Mamy z nimi do czynienia wówczas, gdy przedmiotem wyjaśniania jest pojedyncze zdarzenie. Rozważmy prosty przykład.

Załóżmy, że pytanie o wyjaśnienie ma postać: „Dlaczego Jan wyzdrowiał z infekcji bakteryjnej po zastosowaniu penicyliny?” — „dlaczego  $G(b)$ ?”. Odpowiadając na to pytanie, przyjmujemy, że u Jana wystąpiła infekcja bakteryjna, ponadto musimy przytoczyć twierdzenie statystyczne, stwierdzające wysokie prawdopodobieństwo polepszenia stanu zdrowia pacjentów po zastosowaniu penicyliny w przypadku występowania u nich infekcji bakteryjnej. Schemat takiego typu wyjaśnień posiada zatem postać<sup>9</sup>:

SCHEMAT 1

$$\begin{array}{l} (1) P(G/F)=p \\ (2) F(b) \\ \hline (3) G(b) \end{array} [p]$$

Przesłanka (1) jest twierdzeniem statystycznym stwierdzającym, że względna częstość występowania przedmiotów typu  $F$  wśród przedmiotów typu  $G$  wynosi  $p$ , gdzie wartość  $p$  jest bardzo bliska 1, czyli że prawdopodobieństwo wyzdrowienia z infekcji bakteryjnej po zastosowaniu penicyliny jest bardzo wysokie. Zauważmy, że przesłanka (1) nie jest prawem statystycznym. Dokonane w niej ustalenia przeprowadzone są na podstawie badań biomedycznych na skończonych zbiorach pacjentów, jednak badania te nie wyczerpują zbioru wszystkich możliwych pacjentów poddanych terapiom. Jest to zatem twierdzenie, w którym operujemy pojęciem prawdopodobieństwa *a posteriori*, czyli jest to czasoprzestrzennie zamknięta generalizacja historyczna. Podkreślić należy, że przykład ten jest w pełni reprezentatywny dla badań biomedycznych, w których w procesie wyjaśniania przytaczane są najczęściej statystyczne generalizacje historyczne, a nie prawa statystyczne.

Przesłanka (2) stwierdza, że przedmiot  $b$  posiada cechę  $F$ , czyli, że u Jana występuje infekcja bakteryjna. W (3) zatem wnosimy, z prawdopodobieństwem równym  $p$ , że przedmiot  $b$  posiada również cechę  $G$ . Podwójna linia oddzielająca wniosek od przesłanek wskazuje na to, że mamy do czynienia z wnioskowaniem indukcyjnym, a nie dedukcyjnym.

Oczekiwanie, że wyjaśnienie statystyczne jest poprawne wówczas, gdy prawdopodobieństwo wystąpienia danego zdarzenia jest bardzo wysokie,

<sup>9</sup> Problemy ujęcia indukcyjno-statystycznego podane są za SALMON, 2006, 53–60.

wyduje się racjonalną postawą podmiotu poznającego. Warunek ten jednak jest niestety również źródłem pewnego problemu. Przyjrzyjmy się mu bliżej, rozważając następującą sytuację.

Załóżmy, że rozważamy przypadek terapii pacjentów, u których choroba spowodowana jest szczepami bakterii opornymi na działanie penicyliny. W takich przypadkach szybki powrót do zdrowia po zastosowaniu kuracji penicyliną jest mało prawdopodobny. W sytuacji tej wyjaśnianie statystyczne przebiegałoby zgodnie z następującym schematem wnioskowania:

SCHEMAT 2

$$\begin{array}{l} 1) P(G/F \wedge J) = p^* \\ 2) \frac{F(b) \wedge J(b)}{G(b)} \\ 3) \end{array} [p^*]$$

gdzie  $P(G / F \wedge J)$  to prawdopodobieństwo wystąpienia zdarzenia  $G$  ze względu na zajście zdarzeń  $F$  i  $J$ .  $G$  — to szybki powrót do zdrowia,  $F$  — infekcja bakteryjna,  $J$  — oporność bakterii na zastosowanie penicyliny, a  $p^*$  to wartość prawdopodobieństwa bardzo bliska 0.

Zauważmy, że w obu wnioskowaniach Schemat 1 oraz Schemat 2 mamy wyprowadzony ten sam wniosek na podstawie przesłanek, które są zgodne logicznie. Problem polega na tym, że przesłanki występujące w Schemacie 2 mogą być również użyte dla zbudowania schematu następującego wnioskowania:

SCHEMAT 3

$$\begin{array}{l} 1) P(\neg G/F \wedge J) = 1 - p^* \\ 2) \frac{F(b) \wedge J(b)}{\neg G(b)} \\ 3) \end{array} [1 - p^*]$$

gdzie  $\neg G(b)$  to brak polepszenia stanu zdrowia u osoby  $b$ . Ponieważ  $p^*$  w schemacie 2 było bliskie 0, zatem  $1 - p^*$  jest bliskie 1. Reszta oznaczeń pozostaje bez zmian.

Mamy zatem dwa wnioskowania reprezentowane za pomocą schematów 1 oraz 3, które posiadają logicznie zgodne przesłanki, ale których wnioski są sprzeczne. Sytuacja taka nie mogłaby wystąpić we wnioskowaniach dedukcyjnych. Obowiązuje w nich bowiem prawo:

$$(p \rightarrow q) \vdash (p \wedge r \rightarrow q)$$

W rozumowaniach probabilistycznych niepoprawny jest analogiczny schemat wnioskowania. Jeżeli  $P(G / F)$  jest bardzo bliskie 1, to  $P(G/F \wedge J)$  może mieć wartość dowolnie bliską 0 (SALMON 2006, 55]. W konsekwencji  $P(\neg G/F \wedge J)$  może mieć wartość bliską 1 tak jak  $P(G/F)$ .

Zauważmy, że opisana sytuacja jest kontrintuicyjna wyłącznie wówczas, gdy przyjmujemy perspektywę logiki klasycznej. Akceptując dedukcyjny charakter wyjaśnień formułowanych w nauce, za paradoksalne uznajemy te sytuacje, w których dwa sprzeczne eksplananda  $Z_1$  i  $Z_2$  wyjaśniane są w oparciu o logicznie zgodne przesłanki. Jeżeli jednak odłożymy nasze uprzedzenia, to opisana sytuacja przestaje zadziwiać swoją paradoksalnością. Przedmiotem naszych analiz są przecież wnioskowania statystyczne. Cóż zatem dziwnego w tym, że poszerzenie przesłanki o dodatkową treść (założenie o antybio-tykooporności pewnych bakterii) pozwala nam wyjaśnić sytuację, w której Jan nie wyzdrowiał? Twierdzenia statystyczne opisują wyłącznie prawdopodobieństwo występowania pewnych zjawisk.

Zdaniem W. Salmona wskazany problem można rozwiązać poprzez rozwinięcie alternatywnego wobec modelu I-S ujęcia wyjaśniania statystycznego. Tę właśnie drogę obrał Salmon, rozwijając model istotności statystycznej.

## 5. TEORIA ISTOTNOŚCI STATYSTYCZNEJ (S-R) WESLEYA SALMONA

W koncepcji S-R uznaje się, że eksplanans nie ma wskazywać na wysokie prawdopodobieństwo zdarzenia opisanego w eksplanandum, lecz powinien identyfikować czynniki statystycznie istotne dla wystąpienia tego zdarzenia (SALMON 2006, 59; HAJDUK 2016, 31). Są to zatem te czynniki, których wystąpienie zmienia prawdopodobieństwo zajścia zdarzenia z eksplanandum<sup>10</sup>. Formułując warunek istotności statystycznej, dokonujemy zatem porównania dwóch wartości prawdopodobieństwa: prawdopodobieństwa *wstępnego* zajścia zdarzenia opisanego w eksplanandum i prawdopodobieństwa zajścia tego zdarzenia z uwagi na wystąpienie zdarzenia opisanego w eksplanansie<sup>11</sup>. W celu zilustrowania swojej tezy Salmon rozważa przykład analizowany przez Hempla.

<sup>10</sup> Podkreślić należy: zmienia, a niekoniecznie zwiększa prawdopodobieństwo wstępne (SALMON, 67).

<sup>11</sup> Podkreślić należy, że *prawdopodobieństwa wstępnego* (*prior probability*) nie należy mylić z *prawdopodobieństwem a priori*. Prawdopodobieństwo wstępne to prawdopodobieństwo przy-

Załóżmy, że stwierdzono szybką poprawę stanu zdrowia osoby  $b$  chorej na pewną infekcję bakteryjną. Osoba  $b$  należy zatem do zbioru  $F$  tych osób, u których rozpoznaje się pewną wspólną cechę  $F$ , jaką jest infekcja bakteryjna. Możemy przyjąć, że prawdopodobieństwo szybkiego powrotu do zdrowia osób ze zbioru  $F$  jest bardzo niskie. W tym jednak zbiorze można wyróżnić dwa podzbiory:  $T_1$  — tych osób, które są leczone penicyliną, i  $T_2$  — tych osób, które nie są leczone penicyliną. Dla osób ze zbioru  $T_1$  prawdopodobieństwo szybkiego powrotu do zdrowia jest wysokie. Wyjaśnienie polega zatem na pokazaniu, że prawdopodobieństwo szybkiego powrotu do zdrowia osoby  $b$ , z uwagi na fakt, że osoba ta należy do zbioru  $T_1$  jest wyższe niż prawdopodobieństwo szybkiego powrotu do zdrowia osoby  $b$  wyłącznie na podstawie tego, że jest ona elementem zbioru  $T_2$ . Przyjmijmy, że  $P(G_b/\neg T)$  to prawdopodobieństwo szybkiego wyzdrowienia osoby  $b$ , u której nie zastosowano penicyliny, natomiast  $P(G_b/T)$  to prawdopodobieństwo wyzdrowienia osoby  $b$  z uwagi na zastosowanie penicyliny. Można zatem stwierdzić, że:

$$P(G_b/T) > P(G_b/\neg T)$$

Należy jednak pamiętać, że zbiór  $T_1$  może podlegać dalszym podziałom na zbiór  $J_1$  oraz  $J_2$ .  $J_1$  — zbiór osób, u których infekcja bakteryjna została wywołana szczepem opornym na antybiotyki;  $J_2$  — zbiór osób, u których infekcja została wywołana szczepem podatnym na działanie antybiotyku. Podziały te najczęściej można kontynuować dalej. Spostrzeżenia te stanowią podstawę dla następujących ogólnych ustaleń.

Wyjaśniając statystycznie zajście konkretnego zdarzenia, dokonujemy podziału logicznego uniwersum na podzbiory. Możemy następnie stwierdzić, że dokonany podział jest *istotny* ze względu na pewien atrybut/cechę  $g$ , jeżeli prawdopodobieństwo wystąpienia tego atrybutu w danym podzbiorze jest różne od prawdopodobieństwa jego wystąpienia w innych podzbiorach, będących członami podziału. Ścisłej, jest ono różne dla każdego podzbioru, czyli:  $P(G_b/T) \neq P(G_b/\neg T)$ .

---

porządkowywane zdarzeniu wstępnie. Może mieć charakter subiektywny, związany z personalistyczną interpretacją prawdopodobieństwa, może być ustalany na podstawie doświadczenia wcześniejszego lub może być oparte na interpretacji statystycznej. Tylko w ostatnim przypadku oba pojęcia są znaczeniowo równoważne. Rozróżnienie prawdopodobieństwa wstępnego od *a priori* jest szczególnie istotne w analizach procedury diagnostycznej. Diagnosta przypisuje wstępnie pewną wartość prawdopodobieństwa hipotezom diagnostycznym, ale nie oznacza to, że jest to prawdopodobieństwo *a priori*. Przeciwnie, zwykle *a posteriori* lub subiektywistycznie. Por. RZEPIŃSKI (w druku).

W dalszej kolejności możemy powiedzieć, że zbiór jest *homogeniczny* z uwagi na atrybut  $g$ , jeżeli nie można dokonać w tym zbiorze żadnego podziału, który byłby istotny z uwagi na  $g$ <sup>12</sup>. Można przy tym wyróżnić dwa aspekty. Powiemy mianowicie, że zbiór jest *epistemicznie homogeniczny* wówczas, gdy nie wiemy, w jaki sposób można dokonać istotnego podziału tego zbioru, czyli posiadana wiedza nie dostarcza informacji, jak taki podział przeprowadzić. Zbiór jest natomiast *obiektywnie homogeniczny* wówczas, gdy nie jest możliwe, bez względu na posiadaną wiedzę, dokonanie istotnego podziału zbioru (SALMON 2006, 63).

W tym miejscu nasuwa się pewne zastrzeżenie. Nie jesteśmy przecież w stanie przekonać się w danym czasie, czy określenie homogeniczności zbioru ma charakter obiektywny, czy też jest spowodowane wyłącznie ograniczeniem naszej wiedzy. Zdając sobie sprawę z tego zastrzeżenia, W. Salmon precyzuje pojęcie *homogenicznie istotnego podziału*, nie rozstrzygając, czy ma on charakter epistemiczny, czy obiektywny. Podział jest homogenicznie istotny wówczas, gdy każdy człon tego podziału jest homogeniczny, czyli gdy nie da się dokonać już dalszych podziałów homogenicznie istotnych.

Odnieśmy te ustalenia do analizowanego wcześniej przykładu. Elementami zbioru  $F$  są osoby, u których rozpoznano infekcję bakteryjną, czyli mające cechę  $F$ . Z uwagi na zastosowanie leczenia penicyliną zbiór zawierający elementy posiadające cechę  $F$  został podzielony na dwa podzbiory. Dokonany podział jest istotny, ponieważ odmienne są wartości prawdopodobieństwa dla elementów tych zbiorów z uwagi na atrybut  $G$  — szybki powrót do zdrowia. Następnie można dokonać podziału zbiorów  $T_1$  i  $T_2$  z uwagi na to, czy bakterie, które wywołały infekcję, są odporne na zastosowanie antybiotyku na podzbiory  $J_1$  oraz  $J_2$ . Uzyskujemy zatem następujące klasy podziału:

$\{T_1 \cap J_1\}$  — zbiór osób poddanych terapii penicyliną, u których zakażenie wywołane było bakteriami opornymi na działanie penicyliny;

$\{T_1 \cap J_2\}$  — zbiór osób poddanych terapii penicyliną, u których zakażenie wywołane było bakteriami podatnymi na działanie penicyliny;

$\{T_2 \cap J_1\}$  — zbiór osób nie poddanych terapii penicyliną, u których zakażenie wywołane było bakteriami opornymi na działanie penicyliny;

$\{T_2 \cap J_2\}$  — zbiór osób nie poddanych terapii penicyliną, u których zakażenie wywołane było bakteriami podatnymi na działanie penicyliny.

<sup>12</sup> Czyli w omawianym przykładzie podział zbioru  $F$  na  $T_1$  i  $T_2$  nie był homogeniczny, ponieważ można było później jeszcze dokonać podziału  $T_1$  z uwagi na antybiotykooporność na  $J_1$  i  $J_2$ .

Przyjmijmy, że symbol  $T$  reprezentuje cechę osób ze zbioru  $T_1$ . Odpowiednio  $\neg T$  reprezentuje cechę osób ze zbioru  $T_2$ . Analogicznie  $J$  oraz  $\neg J$  reprezentują cechy ze zbiorów  $J_1$  oraz  $J_2$ .

W wyniku przeprowadzenia podziałów logicznych otrzymujemy cztery człony podziału, ale prawdopodobieństwo wystąpienia atrybutu  $G$  (szybkiego powrotu do zdrowia) w poszczególnych członach podziału nie jest różne. Mamy bowiem do czynienia z sytuacją, w której:  $P(G/T \cap J) = P(G/\neg T \cap J) = P(G/\neg T \cap \neg J)$ . Wyłącznie  $P(G/T \cap \neg J)$  różni się od prawdopodobieństwa wystąpienia atrybutu  $G$  w pozostałych członach podziału (SALMON 2006, 64–65)<sup>13</sup>.

Uogólniając dokonane ustalenia, można powiedzieć, że wyjaśnianie statystyczne w modelu S-R jest poprawne wówczas, gdy spełnione są następujące warunki.

$$1^\circ P(G/F) \neq P(G/F \cap C_i)^{14}$$

$$2^\circ P(G/F \cap C_1) = p_1, P(G/F \cap C_2) = p_2, \dots P(G/F \cap C_n) = p_n$$

gdzie dla dowolnego  $i$  oraz  $j \leq n$ ,  $p_i \neq p_j$ , czyli podziały są homogenicznie istotne (SALMON 2006, 65).

Pozostaje zastanowić się, czy spełnienie tych warunków wystarcza do tego, aby pewne wyjaśnienia formułowane w naukach biomedycznych rozpoznawać jako poprawne wyjaśnienia statystyczne.

## 6. KONCEPCJA ISTOTNOŚCI STATYSTYCZNEJ W WYJAŚNIANIU DANYCH POCHODZĄCYCH Z BADAŃ BIOMEDYCZNYCH

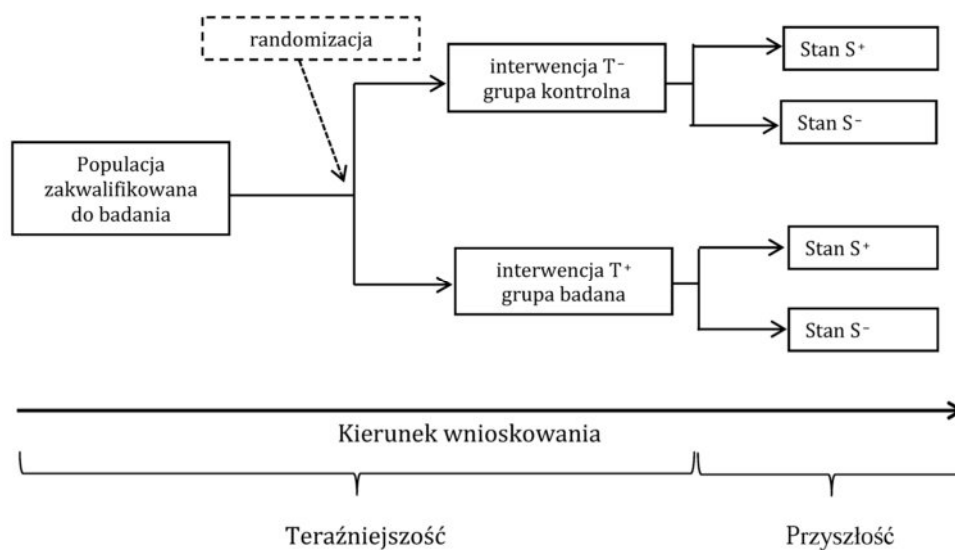
W niniejszym paragrafie rozważone zostaną postulaty koncepcji S-R w kontekście wyjaśnień statystycznych tych danych, które są uzyskane w analizie podgrup randomizowanych badań klinicznych (RCT).

<sup>13</sup> W odniesieniu do omawianego przykładu podziałem istotnym byłoby natomiast wyróżnienie następujących dwóch zbiorów:  $A = \{T_1 \cap J_2\}$  oraz  $B = \{T_1 \cap J_1\} \cup \{T_2 \cap J_1\} \cup \{T_2 \cap J_2\}$ . Prawdopodobieństwa szybkiego powrotu do zdrowia dla pacjentów ze zbiorów  $A$  i  $B$  byłyby bowiem różne.

<sup>14</sup> Jak pisałem wcześniej, prawdopodobieństwo wstępne ma być różne od prawdopodobieństwa w kolejnych członach podziału, niekoniecznie większe. Jak pisze W. Salmon: jeżeli dokonujemy podziału istotnego, to prawdopodobieństwo wystąpienia pewnego zdarzenia w niektórych członach podziału może być większe niż prawdopodobieństwo wstępne, w innych natomiast może być mniejsze (SALMON 2006, 67).

Randomizowane badania kliniczne są rodzajem badań eksperymentalnych z udziałem ludzi, zwykle pacjentów, u których stwierdzono występowanie określonego schorzenia<sup>15</sup>. Idea tych badań jest następująca. W populacji osób zakwalifikowanych do badania wyodrębnia się grupy (kohorty) osób poddanych interwencji oraz grupę kontrolną. Zarówno grupa kontrolna, jak i grupy badane obejmują pacjentów z określonym rozpoznaniem jednostki chorobowej. W grupach badanych testuje się nowe rodzaje interwencji. W grupie kontrolnej stosuje się natomiast placebo lub tradycyjną terapię. Czynniki lub czynniki interwencyjne wprowadzane są w sposób intencjonalny przez badacza. Ważną cechą badania RCT, poza jego eksperymentalnym charakterem, jest losowa alokacja pacjentów do poszczególnych kohort. Przebieg badania RCT ilustruje poniższy schemat.

SCHEMAT 4



T<sup>+</sup> — zastosowanie interwencji, T<sup>-</sup> — niezastosowanie interwencji, S<sup>-</sup> — brak oczekiwanego efektu, S<sup>+</sup> — oczekiwany efekt<sup>16</sup>.

W dalszej części artykułu stan S<sup>-</sup> będzie określany mianem „zdarzenia niekorzystnego”.

W badaniu RCT przeprowadza się wnioskowanie progresywne, czyli wnioskujemy od przyczyny do skutku. Przeprowadzenie badania RCT

<sup>15</sup> Nie zawsze. W fazie 0 i większości badań fazy I wykorzystuje się zdrowych ochotników.

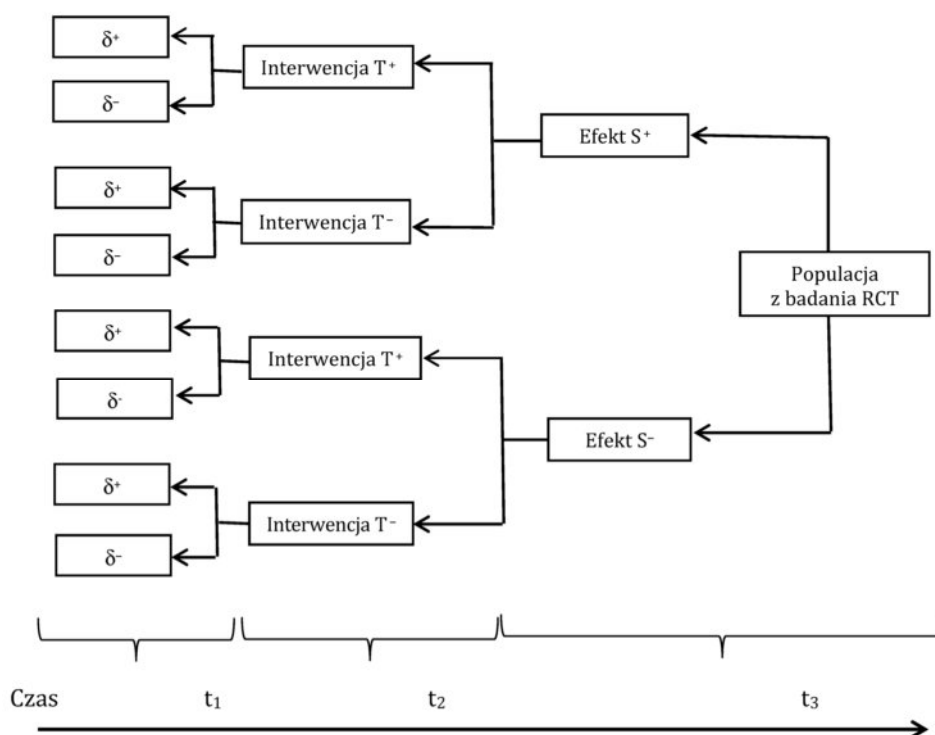
<sup>16</sup> Za: BEAGLEHOLE, BONITA i KLEJSTRÖM 2002, 55 oraz RZEPIŃSKI 2016.



związane jest zatem z dokonaniem podziału populacji osób zakwalifikowanych do badania na podgrupy według atrybutu  $g$  — poddanie terapii. Jeżeli prawdopodobieństwo poprawy stanu zdrowia w poszczególnych grupach będzie różne, to powiemy — zgodnie z ustaleniami W. Salmona — że podział ten jest istotny. Problem polega na tym, że podział ten nie jest homogeniczny. Innymi słowy, badacz wykonujący badanie ma zwykle podstawy, aby sądzić, że w historii pacjenta były takie istotne czynniki, które wpłynęły na efekt terapii.

Badacz przeprowadzający badanie RCT może zatem rozważyć, czy pewna cecha  $\delta^+$  jest istotna dla efektów terapii. Przyjmując tego rodzaju hipotezę, badacz przeprowadza analizę podgrup badania RCT. Sprawdza mianowicie, jaka liczba pacjentów z interwencją  $T^+$  posiada cechę  $\delta^+$ , a jaka liczba posiada cechę  $\delta^-$ . Takie samo ustalenie jest dokonywane wobec pacjentów z grupy kontrolnej bez interwencji. Schemat tego badania ma następującą postać.

SCHEMAT 5<sup>17</sup>



<sup>17</sup> RZEPIŃSKI 2016.

Celem poznawczym analizy podgrup jest zatem zidentyfikowanie takich czynników występujących w historii przyczynowej pacjentów w czasie  $t_1$ , które wpłynęły na przebieg interwencji podjętej w czasie  $t_2$ , dostarczając efektów w czasie  $t_3$  (OXMAN i GUYATT 1992, 79).

Zauważmy, że przy początkowym podziale pacjentów w badaniu RCT na dwie podgrupy ( $T^+$  oraz  $T^-$ ) ostatecznie w analizie podgrup zostają wyróżnione 4 podzbiory dla efektu  $S^+$ . Są to:

$$C_1 = \{\delta^+ \cap T^+\}, C_2 = \{\delta^- \cap T^+\}, C_3 = \{\delta^- \cap T^-\}, C_4 = \{\delta^+ \cap T^-\}$$

Zgodnie z ustaleniami koncepcji S-R możemy uznać, że jeżeli  $P(S^+ / C_i) \neq P(S^+ / C_j)$  dla różnych  $i$  oraz  $j$ , to podział pacjentów, u których zaobserwowano pozytywny efekt terapii  $S^+$  jest podziałem istotnym z uwagi na czynniki  $\delta$  oraz  $T$ . Można zatem uznać, że wyjaśnienie statystyczne pozwala zidentyfikować czynnik, który posiadał wpływ na efekty terapii, prowadząc jednocześnie do polepszenia stanu zdrowia. Podobne ustalenia można dokonać w odniesieniu do  $S^-$ .

Przyjrzyjmy się przykładom zaczerpniętym z nauk biomedycznych. Pierwszy dotyczy badania mającego na celu ocenę skuteczności terapii streptokinazą osób po przebytych zawale serca. W 1986 r. wykonane zostało badanie RCT, mające na celu ocenę skuteczności tej terapii (badanie GISSI)<sup>18</sup>. Uzyskane w badaniu wyniki stały się podstawą do przeprowadzenia analizy podgrup. Jej celem było ustalenie wpływu dodatkowych dwóch czynników, które — jak sądzono — mogą oddziaływać na przebieg terapii. Pierwszym był czas, jaki upłynął u pacjentów biorących udział w badaniu GISSI od początku zawału serca, drugim — wiek pacjentów. Przeprowadzona analiza pozwoliła stwierdzić, że zastosowanie streptokinazy jest bardziej korzystne w grupie osób poniżej 65 roku życia lub wówczas, gdy terapia została podjęta w czasie krótszym niż 6 godzin od pierwszych symptomów zawału (GISSI 1986; LEE, WEISBERG, BRAND i in. 1989). Ponieważ dokonane podziały pacjentów z uwagi na atrybuty: leczenie streptokinazą  $T^+$  oraz wiek  $\delta^+$  były podziałami istotnymi, wyjaśnienie zatem, dlaczego u konkretnej osoby w wieku poniżej 65 roku życia terapia  $T^+$  była skuteczna, jest wyjaśnieniem poprawnym, zgodnie z modelem S-R. Analogicznie można wyjaśnić, dlaczego osoba poddana terapii w wieku powyżej 65 roku życia nie odniosła korzyści.

<sup>18</sup> Streptokinaza to jeden z leków, mających na celu „rozpuszczenie” skrzepliny zamykającej światło tętnicy wieńcowej w trakcie zawału.

Drugi przykład dotyczy analizy podgrup przeprowadzonej na podstawie randomizowanego badania klinicznego ISIS-2. Celem badania była ocena skuteczności prewencyjnego stosowania aspiryny w zapobieganiu zawału serca. Aby ukazać niebezpieczeństwo analizy podgrup, autorzy tego badania wyróżnili podgrupy osób z uwagi na ich przynależność do różnych znaków astrologicznych. Innymi słowy, w zbiorze pacjentów, którzy odnieśli korzyść zdrowotną z prewencyjnego stosowania aspiryny, autorzy badania wyróżnili dwie podgrupy: podzbiór osób urodzonych pod znakiem Bliźniąt lub Wagi oraz podzbiór osób urodzonych pod pozostałymi znakami zodiaku. Okazało się, że prawdopodobieństwo uzyskania korzyści zdrowotnych w grupie pacjentów urodzonych pod znakiem Bliźniąt lub Wagi było znacząco mniejsze niż prawdopodobieństwo uzyskania takich korzyści w grupie pacjentów, którzy urodzili się pod innym znakiem zodiaku (RZEPIŃSKI 2016).

Przyjmijmy że G to szybki powrót do zdrowia, F to prewencyjna terapia aspiryną, a C to cecha urodzenia pod znakiem Bliźniąt lub Wagi. Wówczas:

$$P(G/F \cap \neg C) > P(G/F \cap C)$$

W konsekwencji powinniśmy uznać, że wykorzystując model S-R, możemy poprawnie wyjaśnić, dlaczego osoba urodzona pod znakiem Bliźniąt nie uzyskała korzyści terapeutycznych z leczenia aspiryną<sup>19</sup>. Eksplanans zawierałby twierdzenie wskazujące na to, że prawdopodobieństwo uzyskania korzyści terapeutycznych z prewencyjnego stosowania aspiryny w celu zapobieżenia zawałowi serca jest mniejsze w grupie pacjentów urodzonych pod znakiem Bliźniąt lub Wagi niż prawdopodobieństwo takich korzyści w grupie pacjentów urodzonych pod innymi znakami zodiaku. Wyjaśnienie takie wydaje się jednak dość kontrowersyjne.

## 7. ZAKOŃCZENIE

Współczesna dyskusja nad strukturą wyjaśnień statystycznych przebiega w tradycji ustaleń dokonanych w modelu nomologiczno-dedukcyjnym. Celem rozważań jest określenie poprawności formułowanych wyjaśnień poprzez rozpoznanie struktury eksplanansu. W koncepcji W. Salmona warunkiem poprawności wyjaśnień jest stwierdzenie statystycznie istotnych różnic prawdopodobieństw warunkowych. Praktyka badawcza z zakresu nauk, w których statystyka posiada zastosowanie, wyraźnie jednak pokazuje,

<sup>19</sup> Bardziej szczegółowa analiza tego przykładu w RZEPIŃSKI 2013.

że warunek istotności statystycznej nie wystarcza do określenia prawidłowych struktur eksplanansu wyjaśnień statystycznych. Twierdzenia przytaczane w eksplanansie mogą być rozpoznawane jako błędne, niesatysfakcjonujące, pomimo że identyfikują statystycznie istotne różnice.

Podkreślić należy, że badacze nauk przyrodniczych (przynajmniej pewna ich część) są w pełni świadomi niebezpieczeństw związanych ze stosowaniem kryterium istotności statystycznej jako kryterium określającego poprawność struktury eksplanansu. Doskonale świadczy o tym przytoczony w poprzednim paragrafie przykład analizy podgrup, w której rozważano wpływ znaków zodiaku na efekty prewencyjnego stosowania aspiryny. Jest to przykład tworzenia twierdzeń, które potencjalnie mogą pełnić funkcję eksplanansu, rozpoznawane są jednak jako błędne. Uzyskane wyniki stanowiły dla autorów badania wyłącznie sugestywną ilustrację niebezpieczeństw bezkrytycznego posługiwania się analizą podgrup, w której wiarygodność uzyskanych wyników określana byłaby wyłącznie na podstawie istotności statystycznej.

W badaniach nauk biomedycznych stwierdzenie występowania różnic statystycznie istotnych jest tylko warunkiem koniecznym wnioskowania o zależnościach pomiędzy pewnymi cechami badanych obiektów, nie jest natomiast warunkiem wystarczającym. Poprawność wyjaśnień statystycznych nie zależy zatem wyłącznie od struktury eksplanansu, identyfikującego różnice istotne statystycznie, lecz od wiarygodności procedur badawczych, które stanowiły podstawę dla sformułowania twierdzeń wchodzących w skład eksplanansu. Wyraźnie ilustrują to przytoczone wcześniej przykłady. Wyniki analizy podgrup przeprowadzonej na podstawie badania GISSI nie zostały potwierdzone w późniejszych badaniach randomizowanych. Nie wykazano bowiem żadnych statystycznie istotnych różnic między efektami leczenia pacjentów różniących się czasem podania leku (do 24 godzin) oraz wiekiem (ISIS-2 1988).

W świetle powyższych spostrzeżeń mitem wydaje się przekonanie filozofów nauki, sądzących, że można określić poprawność wyjaśnień statystycznych wyłącznie poprzez analizę struktur twierdzeń wchodzących w skład eksplanansu. Poprawność wyjaśnień statystycznych zależy w zasadniczy sposób od wcześniej zastosowanych metod pozyskiwania danych empirycznych. Podstawą dla tworzenia generalizacji statystycznych mogą być różne typy badań biomedycznych, poczynając od randomizowanych badań klinicznych, poprzez badania kohortowe, badania kliniczno-kontrolne, metaanalizy oraz analizy podgrup. Badania te dostarczają jednak danych

empirycznych o różnym stopniu wiarygodności. Twierdzenia statystyczne formułowane na ich podstawie nie różnią się wprawdzie strukturą, lecz tym, jakie funkcje poznawcze są im przypisywane. Generalizacje statystyczne, identyfikujące statystycznie istotne różnice na podstawie analizy podgrup, są traktowane jako hipotezy inicjujące dalsze badania. Proces ich konstruowania jest przedmiotem analiz w kontekście odkrycia. Nie są one natomiast, względnie też nie powinny być, podstawą dla formułowania wyjaśnień zjawisk. W konsekwencji nie mogą być one również punktem wyjścia do formułowania algorytmów decyzyjnych praktyki medycznej. Te ostatnie tworzone mogą być wyłącznie na podstawie wyjaśnień statystycznych formułowanych w oparciu o bardziej wiarygodne metody badawcze, jak na przykład badania RCT.

#### BIBLIOGRAFIA

- AJDUKIEWICZ, Kazimierz. 1975. *Logika pragmatyczna*. Wyd. III. Warszawa: PWN.
- GILLIES, Donald. 2010. *Philosophical Theories of Probability*. London, New York: Routledge, Taylor & Francis Group.
- GISSI. 1986. „Effectiveness of intravenous thrombolytic treatment in acute myocardial infarction. Gruppo Italiano per lo Studio della Streptochinasi nell'Infarto Miocardico (GISSI)”. *The Lancet* 2: 397–402.
- HAJDUK, Zygmunt. 2016. *Struktury metodologiczne w nauce. Słowa klucze filozofii nauki*. Lublin: Wydawnictwo KUL.
- HEMPEL, Carl G., i Paul OPPENHEIM. 1948. „Studies in the Logic of Explanation”. *Philosophy of Science*, 15: 567–579.
- HEMPEL, Carl G. 1965. *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*. New York: The Free Press
- ISSIS-2 [Second International Study of Infarct Survival]. 1988. „Randomised trial of intravenous streptokinase, oral aspirin, both, or neither among 17 187 cases of suspected acute myocardial infarction”. *Lancet* 2: 349–360.
- KMITA, Jerzy. 1975. *Wykłady z logiki i metodologii nauk*. Warszawa: PWN.
- LEE, Thomas H., Monica C. WEISBERG, Donald B. BRAND, Gregory W. ROUAN i Lee GOLDMAN. 1989. „Candidates for thrombolysis among emergency room patients with acute chest pain”. *Annals of Internal Medicine*, 110 (12): 957–962.
- OXMAN, Andrew D., i Gordon GUYATT. 1992. „A Consumer's Guide to Subgroup Analyses”. *Annals of Internal Medicine* 116 (1): 78–84.
- RZEPIŃSKI, Tomasz. 2011. „Proces wyjaśniania a rewizja wiedzy na przykładzie odkrycia zjawiska kardioprotekcji”. *Roczniki Filozoficzne* 59 (2): 277–299.
- RZEPIŃSKI, Tomasz. 2016. „Problemy epistemologiczne analizy podgrup randomizowanych badań klinicznych”. *Filozofia Nauki* 24 (4): 73–94.
- RZEPIŃSKI, Tomasz (w druku). „Twierdzenie Bayesa w projektowaniu strategii diagnostycznych w medycynie”. *Diametros*.

- SALMON, Wesley C. 2006. *Four Decades of Scientific Explanation*. Pittsburgh, Pa.: University of Pittsburgh Press.
- YUSUF, Salim, Janet Wittes, Jeffrey PROBSTFIELD i Herman A. TYLOR. 1991. „Analysis and Interpretation of Treatment Effects in Subgroups of Patients in Randomized Clinical Trials”. *Journal of American Medical Association* 266 (1): 93-98.

STRUKTURY WYJAŚNIANIA STATYSTYCZNEGO  
WARUNEK ISTOTNOŚCI STATYSTYCZNEJ  
W ANALIZIE PODGRUP RANDOMIZOWANYCH BADAŃ KLINICZNYCH

Streszczenie

Przedmiotem analizy podjętej w artykule jest procedura wyjaśniania statystycznego. Dla potrzeb prowadzonych rozważań omówiony został szczególny typ badań biomedycznych, jakim jest analiza podgrup randomizowanych badań klinicznych. Pokazane zostało, że poprawności wyjaśnień statystycznych nie można ustalić, analizując wyłącznie strukturę eksplanansu, jak jest to przyjmowane w modelu istotności statystycznej (R-S), zaproponowanym przez W. Salmona. Procedura wyjaśniania statystycznego jest zatem, w przeciwieństwie do pozostałych form wyjaśniania, zależna od procedur badawczych stanowiących podstawę formułowania twierdzeń statystycznych.

THE STRUCTURES OF STATISTICAL EXPLANATION:  
STATISTICAL RELEVANCE CONDITION IN SUBGROUP ANALYSIS  
OF RANDOMIZED CLINICAL TRIALS

Summary

The subject of analysis undertaken in the article is the procedure of statistical explanation. For this purpose a specific type of biomedical study, so called subgroup analysis of randomized clinical trials (SA), was discussed. It was shown that the validity of statistical explanations can not be determined by analyzing only the structure of explanation, as is the case in the statistical significance model (S-R) proposed by W. Salmon. Therefore, the statistical explanatory procedure is, in contrast to the other explanatory forms, dependent on the research procedures underlying the formulation of statistical theorems.

**Słowa kluczowe:** wyjaśnianie statystyczne; warunek istotności statystycznej; badania biomedyczne; analiza podgrup.

**Key words:** statistical explanation; statistical relevance condition; clinical studiem; subgroup analysis.

**Information about Author:** Dr. hab. TOMASZ RZEPIŃSKI, prof. at UAM — Department of Logic and Methodology of Sciences, Adam Mickiewicz University Poznan, and Department of Biology and Environmental Protection, Poznan University of Medical Sciences; address for correspondence: ul. Szamarzewskiego 89c, 60-568 Poznań; e-mail: rzeto@amu.edu.pl