

Rozdział 4

Schemat ewaluacji

Krok 2. Określenie schematu ewaluacji

Schemat ewaluacji pozwala Ci odpowiedzieć na pytanie: „Z czym porównywać?” Aby wiedzieć, czy Twój program spowodował jakieś zmiany, musisz dysponować jakimiś wskaźnikami porównawczymi.

Czy zapobiegłeś nadużywaniu substancji uzależniających? Odpowiedź na to pytanie zależy od tego, jak zdefiniowałeś swój wskaźnik sukcesu. Czy stwierdziłbyś, że pewien program profilaktyczny dla szkół prowadzony w USA zakończył się sukcesem, jeśli 5% uczniów starszych klas szkół średnich z rejonu objętego tym programem paliło codziennie marihuanę? A gdybyś wiedział, że kilka lat wcześniej w USA — według ogólnych statystyk — 11% uczniów starszych klas szkół średnich paliło marihuanę codziennie, czy oceniłbyś program jako udany? A wiedząc, że w roku, w którym badano skuteczność programu, ogólne statystyki w USA wykazały, że 4.9% uczniów starszych klas szkół średnich codziennie paliło marihuanę, czy też byś uznał, że program odniósł sukces?

Wszystkie takie porównania są możliwe. Gdybyś nie miał standardów porównawczych, skąd mógłbyś wiedzieć, czy robisz postępy w profilaktyce nadużywania substancji uzależniających? Uzyskane wyniki musisz z czymś porównać, aby ocenić, czy Twoje profilaktyczne działania są skuteczne.

Najbardziej oczywistym i właściwym standardem porównawczym jest porównanie poziomu nadużywania środków uzależniających po zastosowaniu działań profilaktycznych z poziomem tego zjawiska, gdyby takich działań nie przeprowadzano. Zawiera się to w pytaniu: „Jak by się rzecz miała bez naszej interwencji?”

Na przykład, jeśli w ósmych klasach wprowadzasz program profilaktyki antyalkoholowej i pod koniec roku okazuje się, że uczniowie, którzy uczestniczyli w tym programie, pili średnio dwa razy w tygodniu, to odniosłeś sukces czy porażkę? Trudno powiedzieć. Może się wydawać, że picie dwa razy w tygodniu to wysoki wskaźnik. Jednak pedagog szkolny mógł umieścić wszystkich uczniów z pro-

blemami alkoholowymi w klasach objętych Twoim programem wiedząc, że otrzymają specjalną pomoc. Może uczniowie, którzy mówili po Twoim programie, że piją dwa razy w tygodniu, na początku roku szkolnego pili codziennie. Bez danych porównawczych pochodzących z pretestu (określającego średni poziom picia przed wprowadzeniem Twojego programu), nie mógłbyś dowiedzieć się, że spożycie alkoholu spadło o 60% po roku stosowania tego programu. Czy jeśli masz dane z pretestu, możesz stwierdzić, że Twoje działania „spowodowały” 60% ograniczenie używania alkoholu przez uczniów? Nie, dopóki nie przeprowadzisz innych porównań. Mogło się przecież zdarzyć, że wielu ośmioklasistów w szkole na początku roku szkolnego dużo i często piło alkohol, ale związane z nadużywaniem alkoholu śmiertelne wypadki drogowe wśród nastolatków podczas ferii zimowych doprowadziły do 60% redukcji spożycia alkoholu wśród uczniów.

Osoby zajmujące się ewaluacją nazywają to efektem historii zdarzeń — pewnych wydarzeń w określonym czasie, które zmieniają pewne aspekty rzeczywistości. Innym przykładem efektu historii zdarzeń może być wprowadzenie kontroli uczniowskich szafek szkolnych lub procedury komputerowej analizy moczu uczniów podejrzanych o używanie środków uzależniających. Takie wydarzenia też mogą wpłynąć na obniżenie stopnia używania środków uzależniających w szkole. To są możliwe wyjaśnienia obserwowanych wyników, nie związane z przeprowadzonym przez Ciebie programem profilaktycznym.

Poniżej przedstawiono siedem zagrożeń trafności badania opracowanych przez Campbella i Stanleya w 1963 roku.

Siedem zagrożeń trafności badania

HISTORIA ZDARZEŃ: Dotyczy niezaplanowanych wydarzeń mających miejsce pomiędzy Twoim pierwszym w czasie pomiarem (pretestem) a następnym (posttestem). Im dłuższego czasu dotyczy Twoja ewaluacja, tym bardziej musisz zdawać sobie sprawę z możliwości wystąpienia efektu historii zdarzeń.

DOJRZEWANIE: Dojrzewanie odnosi się do zmian rozwojowych, które w naturalny sposób następują wśród osób objętych badaniem. Wraz z wiekiem zdobywają oni więcej doświadczenia i większą niezależność, co również może wpływać na ich zachowanie niezależnie od Twoich oddziaływań profilaktycznych. Jeśli, na przykład, mierzysz zmiany postawy u uczniów ósmoklasistów szkoły podstawowej, musisz pamiętać że są dowody na to, iż postawy społeczne uczniów klas siódmych w sposób naturalny stają się bardziej negatywne w trakcie roku szkolnego.

TESTOWANIE: Efekt testowy dotyczy powtarzania pomiarów tym samym testem na tej samej grupie. Jeśli stosujesz pre- i posttest, musisz być świadom

wystąpienia efektu testowego, czyli oddziaływania testu na osoby badane (np. uczą się one tego, co dla Ciebie jest ważne i spełniają Twoje oczekiwania).

INSTRUMENTACJA: Efekt instrumentalny występuje wówczas, gdy w trakcie ewaluacji wprowadzasz zmiany w narzędziach pomiarowych lub gdy zmieniają się obserwatorzy lub oznaczenia, albo kryteria klasyfikacji zachowań.

REGRESJA STATYSTYCZNA: Regresja statystyczna pojawia się, gdy grupy zostały dobrane na podstawie skrajnych wyników. Przy powtórnych pomiarze ekstremalne wyniki mają tendencję do przesuwania się w kierunku średniego wyniku całej grupy. Jeśli decydujesz się objąć swoim programem uczniów, którzy w preteście wykazywali postawy stosunkowo najbardziej aprobujące używanie środków uzależniających, nie zapominaj, iż jest wielce prawdopodobne, że te wyniki przy kolejnym pomiarze będą mniej ekstremalne nawet bez zastosowania działań profilaktycznych.

SELEKCJA: Zawsze, gdy osoby dobrane do grupy profilaktycznej różnią się od tych z grupy porównawczej/kontrolnej, powstają odchylenia selekcyjne w procesie ewaluacji. Dobór losowy jest najlepszym zabezpieczeniem przed efektem selekcji.

WYKRUSZANIE SIĘ GRUPY: Jeśli występuje nierówne lub selektywne wykruszanie się jednej z grup — profilaktycznej lub porównawczej/kontrolnej — może to mieć wpływ na wartość Twoich wyników. Gdyby Twoje badania dotyczyły efektów programu profilaktycznego dla rodziców mającego za cel zmianę postawy rodzin wobec stosowania środków uzależniających i rodzice uczniów już stosujących te środki — uznając zajęcia za zbyt zagrażające — opuściliby je, to takie wykruszenie zagraża wartości wyników Twoich badań.

Aby, pomimo wpływu różnych losowych zdarzeń i innych czynników, otrzymać jak najlepszy obraz tego, jakby się rzeczy miały bez Twojego programu, powinieneś — dla celu porównania — mieć odpowiednio dobraną grupę nie objętą oddziaływaniami. Grupa poddana oddziaływaniom i nie poddana powinny być do siebie bardzo podobne; najlepiej, gdy stosuje się losowy dobór osób do jednej i drugiej grupy. Podjęcie decyzji, z czym należy porównać wyniki programu, jest równoznaczne z określeniem schematu badania.

Schemat ewaluacji pozwala Ci porównywać wyniki z: (1) poprzednimi osiągnięciami; (2) wynikami pochodzącymi z grup o podobnych charakterystykach, ale poddanych innym działaniom profilaktycznym, lub nie objętych żadnym programem; (3) wynikami pochodzącymi z grup różniących się od Twojej grupy profilaktycznej pod innymi względami. Im „mocniejszy” schemat ewaluacji (im większe jest podobieństwo między grupą profilaktyczną i porównawczą pod względem charakterystyki i doświadczeń nie związanych z progra-

mem), tym jest bardziej prawdopodobne, że będziesz mógł stwierdzić, czy Twój program spowodował jakieś zmiany. Na dalszych stronach zostaną omówione podstawowe schematy badawcze.

Część 1. Rozważ możliwe schematy ewaluacji

Oto cztery schematy najczęściej stosowane w ewaluacji programów profilaktycznych:

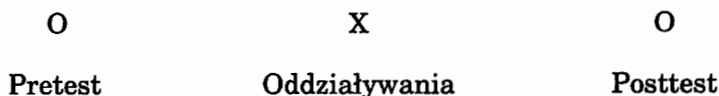
1. Schematy „pretest-posttest”
2. Schematy eksperymentalne
3. Schematy quasi-eksperymentalne
 - a) z grupą porównawczą
 - b) szeregów czasowych
4. Schematy tylko z posttestem

Schematy „pretest-posttest”

Schemat pretest-posttest polega na porównaniu wyników dwóch pomiarów, przed i po zastosowaniu Twojego programu, dokonanych na tych samych grupach osób. Schemat ten nie jest mocnym testem skuteczności programu, ponieważ, na przykład, nie bierze pod uwagę pewnych wydarzeń nie związanych z programem, które mogły wpłynąć na zmianę poziomu używania środków uzależniających w całej populacji. Może jednak wykazać, czy uczestnicy Twojego programu zmienili się w pożądanym kierunku (tj. czy po przeprowadzeniu Twojego programu osiągnąłeś „bezpośrednie wyniki”, jakich oczekiwałeś). Gdy masz, na przykład, przygotowany program zapobiegania używaniu substancji uzależniających i nie jesteś pewien, czy jest on lepiej dostosowany do potrzeb uczniów klas trzecich czy czwartych szkoły podstawowej, możesz zastosować pretest w obu klasach, wprowadzić swój program, a następnie przeprowadzić posttest. Przez porównanie zmian — powstałych w obu klasach między pre- i posttestem — w poziomie wiedzy na temat substancji uzależniających, postawach i intencjach dotyczących używania środków uzyskasz pewne przesłanki ułatwiające podjęcie decyzji, do jakiego poziomu wiekowego adresować Twój program.

Stosując schemat pretest-posttest nigdy nie wiesz, czy zmiany, jakie obserwujesz, zostały wywołane przez Twój program. W porównaniach korzystasz z informacji o tym, jacy byli uczestnicy Twojego programu, zanim go zastosowałeś. Taki schemat mówi Ci, na ile badane przez Ciebie osoby zmieniły się pod pewnymi względami podczas prowadzonych działań. Nie dowiadujesz się jednak, co spowodowało zmiany, ani czy osoby nie objęte Twoim programem również się zmieniły.

Oto diagram schematu pretest-posttest:



Wprowadzenie grup kontrolnych i porównawczych

Rozwiązaniem, które pozwoli Ci sprawdzić, czy to Twój program spowodował zmiany postaw lub zachowań, może być wprowadzenie grup porównawczych lub kontrolnych, jak najbardziej podobnych do grup uczestników Twojego programu pod każdym istotnym względem.

Schematy eksperymentalne

Ponieważ przed wprowadzeniem programu często trudno jest przewidzieć, jakie zmienne mogą mieć zasadnicze znaczenie dla zachowania podobieństwa grup, powszechną metodą tworzenia grup do porównań jest dobór losowy. O tym, kto znajdzie się w grupie eksperymentalnej a kto w kontrolnej, decydować może rzut monetą albo z góry ustalony numer na wspólnej liście. Każdy poprawny schemat eksperymentalny korzysta z jakiegoś konkretnego sposobu losowego doboru osób do grup. Dobór losowy to najlepsza metoda neutralizowania wszelkich różnic pomiędzy osobami objętymi Twoimi programem profilaktycznym a tymi, z którymi będziesz je porównywał. Dzięki temu zdobywasz wysoki stopień pewności, że różnice między grupami występujące po programie zostały wywołane właśnie Twoimi działaniami. Grupa osób nie objętych Twoimi oddziaływaniami zwana jest grupą kontrolną.

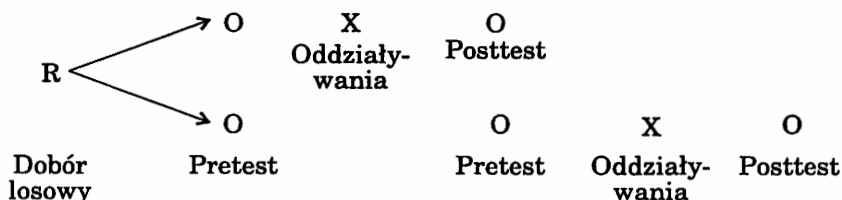
Oto przykład badania z losowym podziałem na grupy. Porozumienie na Rzecz Profilaktyki w Quinn przygotowało dwa programy promujące zdrowy styl życia wśród młodzieży i zniechęcające do eksperymentowania z substancjami uzależniającymi. Pierwszy polegał na przeprowadzeniu przez nauczyciela dziesięciu godzin zajęć edukacyjnych w ciągu pięciu tygodni. W drugim programie — o nazwie „Super Gwiazdy” — w okresie pięciu tygodni odbywało się dziesięć jednogodzinnych spotkań z publicznie znanymi osobami mogącymi służyć młodzieży jako pozytywne wzorce (sportowcy i inni). Chciano się dowiedzieć, który z programów miał większy wpływ na postawy i zachowania młodzieży.

Uczniowie drugiego rocznika szkoły średniej w Waszyngtonie zostali losowo przydzieleni do trzech klas o profilu zdrowotnym, które prowadziła ta sama nauczycielka. We wszystkich trzech klasach przeprowadzono pretest. W klasie numer 1 nauczycielka zastosowa-

ła nowy program zdrowotny. W klasie numer 2 odbyło się dziesięć jednogodzinnych prezentacji „Super Gwiazd”. W klasie numer 3, stanowiącej grupę kontrolną, nauczycielka prowadziła swoje rutynowe lekcje na temat zdrowia. Po upływie pięciu tygodni wszyscy uczniowie wzięli udział w postteście, aby między poszczególnymi klasami porównać poziom aprobowania przez uczniów używania substancji uzależniających oraz poziom deklarowanego przez nich używania tych środków. Dzięki temu można było ocenić, który z programów spowodował bardziej pozytywne zmiany w postawach i zachowaniach uczniów. Oto schemat tego badania:



Drugim przykładem schematu eksperymentalnego jest schemat z odroczonymi oddziaływaniami. W tym modelu wszyscy z Twojej próby badawczej biorą udział w programie profilaktycznym, ale w innym czasie. Stosując schemat z odroczonymi oddziaływaniami losowo decydujesz, kto pierwszy weźmie udział w zajęciach profilaktycznych, a kto w dalszej kolejności. Jedną z trudności związanych z tym schematem jest zadbanie o to, aby osoby, które wcześniej brały udział w programie, nie wpływały na postawy i zachowania tych, którzy dopiero będą w nim uczestniczyć. Taki schemat może być przydatny w sytuacji, gdy masz niewielki zespół osób prowadzących zajęcia i z tego powodu oddziaływania w różnych grupach lub klasach muszą być rozłożone w czasie. Oto diagram schematu z oddziaływaniami odroczonymi w czasie:



Dobór losowy nie zawsze jest łatwą sprawą. Gdy zajmujesz się profilaktyką w szkołach czy w instytucjach użyteczności publicznej, w praktyce może być Ci trudno pracować tylko z częścią populacji,

odmawiając pomocy pozostałym osobom, które w równym stopniu jej potrzebują. Na pierwszy rzut oka dobór losowy może wydawać się niesprawiedliwy — odcina część populacji od wartościowych działań profilaktycznych. Z drugiej strony, jeśli nie masz dość środków, aby zająć się wszystkimi naraz, dobór losowy — rzut monetą — może być najbardziej sprawiedliwym sposobem podjęcia decyzji, kogo objąć działaniami profilaktycznymi, zwłaszcza gdy jest na nie duże zapotrzebowanie. Ponadto do chwili, gdy nie zostanie udowodnione, że Twoje działania są skuteczne, nie jest jasne, czy dobór losowy pozbawia kogoś wartościowych usług. Zastosowanie schematów badawczych z doбором losowym, kiedy realizuje się samodzielną ewaluację programów profilaktycznych może być jednak ograniczane przez priorytety organizacyjne, polityczne i społeczne. W sytuacjach, w których dobór losowy jest niemożliwy, można tworzyć grupy podobne do siebie pod wieloma, choć nie wszystkimi względami. Takie nielosowo dobrane grupy nazywa się grupami porównawczymi. Schematy, w których stosuje się grupy porównawcze, są to schematy quasi-eksperymentalne.

Schematy quasi-eksperymentalne

Schematy z grupą porównawczą. Przypomnij sobie opisany wcześniej przykład z Quinn, gdzie sprawdzano działanie dwóch programów. Przypuśćmy, że uczniowie już wcześniej zostali przydzieleni na ten rok szkolny do klas o profilu zdrowotnym, co uniemożliwia dobór losowy. W takim przypadku można rozważyć zastosowanie kilku schematów quasi-eksperymentalnych. Koordynator badań może zdecydować się na porównywanie efektów programu „Super Gwiazdy” z rezultatami nowego zdrowotnego programu edukacyjnego prowadzonego przez nauczycielkę, nie stosując przy tym losowego doboru do grup. Może też w trzech grupach wprowadzić programy pięcioletniowe: nowe edukacyjne zajęcia zdrowotne w klasie numer 1, dziesięć godzin wizyt „Super Gwiazd” w klasie numer 2 i stary program zdrowotny w klasie numer 3. Na koniec może porównać pomiędzy trzema grupami ich średnie wyniki dotyczące postawy wobec używania środków uzależniających i średnie wyniki deklarowanego przez uczniów używania tych środków. Taki schemat wygląda następująco:

X_1	O
X_2	O
K	O
Oddziaływania/ Kontrola	Posttest

Powyższy schemat można wzmocnić porównując trzy klasy zarówno przed, jak i po zastosowaniu działań profilaktycznych. W takim przypadku będzie to schemat następujący:

O	X ₁	O
O	X ₂	O
O	K	O
Pretest	Oddziaływania/ Kontrola	Posttest

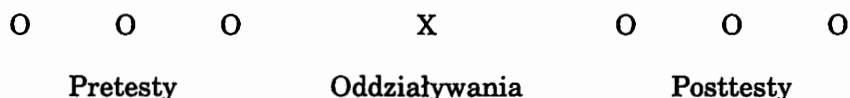
Ten schemat jest bardzo podobny do wcześniej omówionego schematu eksperymentalnego. Nie zastosowano w nim jednak doboru losowego, który gwarantuje, że klasy mają podobny skład.

W Stanach Zjednoczonych istnieje możliwość posłużenia się schematem quasi-eksperymentalnym również w takiej sytuacji, gdy nie ma grupy porównawczej. Można wówczas porównać wyniki programu profilaktycznego z wynikami ogólnokrajowych badań dotyczących używania środków uzależniających przez młodzież. Roczne wyniki tych badań są dostępne w raportach Narodowego Instytutu Badań nad Nadużywaniem Substancji Uzależniających (NIDA — National Institute on Drug Abuse). Raporty te zawierają krajowe i regionalne oszacowania przypadków nadużywania substancji uzależniających i rozpowszechnienia określonych postaw i zachowań związanych z takimi substancjami dla reprezentatywnej próby młodzieży. Oszacowania te mogą być wykorzystywane dla celów porównawczych. Przypuśćmy, że szuka się sposobu zmniejszenia poziomu używania marihuany przez uczniów szkół średnich, wprowadzając program z udziałem uczniów jako asystentów. Jeśli porównuje się poziom spożycia marihuany w szkole objętej programem z poziomem określonym w dorocznych badaniach dotyczących uczniów szkół ponadpodstawowych, stosuje się właśnie schemat quasi-eksperymentalny.

Różnorodność quasi-eksperymentalnych schematów porównywania grup jest nieskończona. W takich schematach możesz stosować dowolną kombinację pomiarów, zwiększając lub zmniejszając ilość obserwacji (O) i działań profilaktycznych (X). Ale pamiętaj, że im więcej obserwacji i działań profilaktycznych obejmuje Twój schemat badawczy, tym bardziej skomplikowany będzie Twój projekt ewaluacji.

Schematy szeregów czasowych. Chociaż zawsze korzystniej jest stosować grupę porównawczą lub kontrolną, może się to okazać niemożliwe. W takim przypadku można polecać schemat szeregów czasowych. Taki schemat jest silniejszy niż schemat typu pretest-posttest, ponieważ pozwala Ci obserwować kierunek zmian zachowania

dzących w zachowaniach i postawach przed i po zastosowaniu Twojego programu. Najpoważniejszym ograniczeniem tego schematu jest trudność wynikająca z konieczności wielokrotnego badania grupy, zanim rozpoczniesz na niej działania profilaktyczne. Oto diagram tego schematu:



Schematy tylko z posttestem

Jeśli zależy Ci na jakimś pomiarze skuteczności swojego programu, ale nie masz dostatecznej liczby uczestników, aby stworzyć grupę kontrolną czy porównawczą, albo nie masz możliwości przeprowadzenia pretestu, możesz uzyskać pewne informacje stosując schemat tylko z posttestem. W takim schemacie przeprowadzasz po prostu badania grupy objętej programem profilaktycznym w jakimś czasie po zakończeniu oddziaływań.

Schemat tylko z posttestem może być zastosowany w sytuacji, gdy z osobami badanymi lub grupą badaną masz kontakt tylko raz i nie jesteś w stanie przeprowadzić dwóch pomiarów: pretestu i posttestu. Można się tu posługiwać przykładem ewaluacji szkolenia rodziców poprzez jednorazowy warsztat prezentowany w wieczornym programie telewizyjnym w jednym ze stanów w USA. Warsztat emitowany był dwa dni po dokumentalnym programie telewizyjnym poświęconym doświadczeniom pewnej rodziny z nastolatkiem uzależnionym od narkotyków. Rodzice, którzy wzięli udział w warsztacie (ponad 4000 jednego wieczoru), byli zachęceni do tego przez zespół do spraw kontaktów z widzami, w trakcie oglądania programu telewizyjnego o rodzinie z uzależnionym dzieckiem i poprzez artykuły w prasie. Ci ludzie, w momencie decydowania się na udział w warsztacie dla rodziców na temat zapobiegania używaniu środków odurzających przez młodzież, nie byli przygotowani do udziału w preteście. Przeprowadzono jednak ewaluację. Zastosowano jednostronicowy posttest dla osób, które oglądały warsztat profilaktyczny i wykorzystano rozmowy telefoniczne słuchaczy z osobami, które koordynowały i prowadziły sesje warsztatowe. Oto diagram schematu tylko z posttestem:

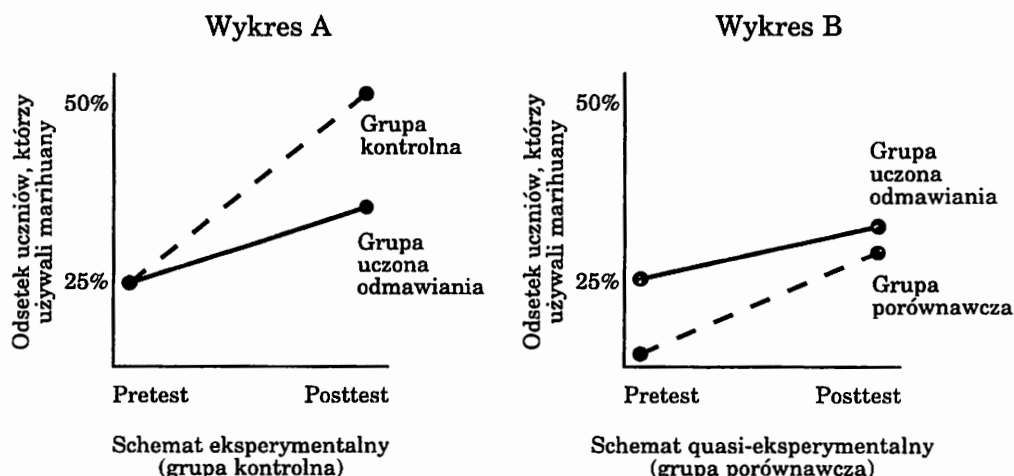


Schematy tylko z posttestem są najłabszymi z omawianych tu schematów. Mogą jednak dostarczać informacji, które ulepszą Twój program pod wieloma względami, dotyczącymi na przykład sprecyzowania obiektu Twoich oddziaływań, określenia społecznego zapotrzebowania na Twój program, sposobu rekrutacji uczestników, zastosowanych metod i treści. Chociaż takie schematy nie pozwalają na porównanie efektów Twojego programu z żadnymi standardami czy grupami albo z wynikami pretestu, mogą być użyte do pomiaru poziomu intencji i używania środków uzależniających po wprowadzeniu programu profilaktycznego.

Porównywanie wyników badań z zastosowaniem różnych schematów

Różnice między tym, czego możesz się dowiedzieć na temat efektów swojego programu stosując któryś ze schematów, można łatwiej zrozumieć porównując graficzne przedstawienia wyników uzyskanych w badaniach według czterech, wyżej omówionych modeli. Przyjrzyj się wykresom A, B, C, i D na rycinie 5.

Te cztery wykresy ilustrują efekty programu profilaktycznego przeznaczonego dla rodziców i nauczycieli. Ma on przygotowywać ich do uczenia dzieci umiejętności odmawiania. Miarą skuteczności programu są tu proporcje uczniów, którzy używali marihuany pod koniec ósmej klasy — rok po przeprowadzeniu programu profilaktycznego.



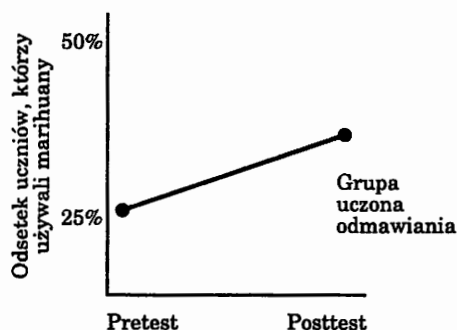
Ryc. 5. Efekty szkolnego i rodzinnego treningu umiejętności odmawiania

Wykres A przedstawia wyniki schematu eksperymentalnego. Według założeń tego schematu uczniowie zostali losowo podzieleni na klasy, w których miał być wprowadzany trening odmawiania i w których nie miało być tego programu. Zauważ, że na początku siódmej klasy (pretest) około 25% uczniów w grupie eksperymentalnej i w grupie kontrolnej już zaczęło palić marihuanę. Pod koniec ósmej klasy wśród uczniów, którzy przeszli trening odmawiania, liczba tych, którzy używali marihuany, wzrosła do około 35%, podczas gdy w grupie kontrolnej w tej samej szkole — aż do 51%. Osoby badające efekty profilaktyki powiedziałyby, że w wyniku treningu odmawiania nastąpił 16-procentowy spadek rozpoczynania eksperymentowania z marihuaną. Ponieważ uczniowie byli losowo podzieleni na klasy i w preteście prezentowali podobny poziom pierwszych prób palenia marihuany, pracownicy profilaktyki mogą być pewni, że różnica, która wystąpiła, jest wynikiem programu uczenia umiejętności odmawiania.

Wykres B przedstawia wyniki badania przy zastosowaniu schematu quasi-eksperymentalnego. W tym badaniu wszyscy uczniowie klas siódmych z jednej szkoły brali udział w szkoleniu w zakresie umiejętności odmawiania, podczas gdy uczniowie klas siódmych z innej szkoły, którzy nie zostali przeszkoleni, stanowili grupę porównawczą. Jak pokazuje wykres B, na początku programu więcej uczniów ze szkoły, w której zaplanowano wprowadzać program profilaktyczny, zaczęło już palić marihuanę (około 25% z całej populacji rozpoczynającej siódmą klasę). Wydaje się, że była to szkoła, w której działania profilaktyczne były wyraźnie potrzebne. W szkole porównawczej tylko około 5% rozpoczynających siódmą klasę paliło marihuanę. Jak widać na wykresie B, wzrost proporcji uczniów rozpoczynających palenie marihuany pod koniec ewaluowanego okresu był większy w szkole, w której nie prowadzono programu rozwijającego umiejętności odmawiania, niż w szkole, w której odbywały się takie zajęcia. Ten wynik jest obiecujący. W grupie profilaktycznej nastąpił jedynie niewielki wzrost poziomu eksperymentowania z marihuaną po przeprowadzeniu treningu umiejętności odmawiania. Ale na podstawie wykresu B nie można określić, co było przyczyną wystąpienia różnic w poziomie eksperymentowania z marihuaną w obu szkołach. Mógł spowodować to trening umiejętności odmawiania. Z drugiej strony mogło to wynikać z różnego poziomu palenia marihuany w obu szkołach na początku siódmej klasy. Może uczniowie w szkole porównawczej pod koniec ósmej klasy zaczęli po prostu nadrobić zaległości w stosunku do uczniów ze szkoły objętej profilaktyką, którzy zaczęli palić wcześniej. Chociaż wyniki wyglądają obiecująco, prowadzący działania profilaktyczne nie mogą być pewni, że to ich program spowodował różnice w poziomie wzrostu palenia marihuany.

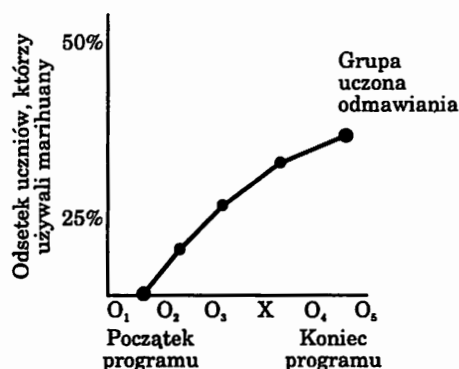
Wykres C przedstawia wyniki badania typu pretest-posttest bez grupy porównawczej. W tym przypadku trening umiejętności odma-

Wykres C



Schemat pretest-posttest

Wykres D



Schemat szeregów czasowych

Ryc. 5. Efekty szkolnego i rodzinnego treningu umiejętności odmawiania, cd.

wiania był wprowadzony w klasach siódmych w jednej szkole, ale nie badano grupy porównawczej. Wykres C pokazuje, że odsetek uczniów, którzy rozpoczęli palenie marihuany, wzrósł po przeprowadzeniu treningu umiejętności odmawiania z 25% do około 35%. Jest to taki sam poziom wzrostu jak w poprzednim przykładzie. Czy więc ten program zakończył się sukcesem, czy porażką? Poziom palenia marihuany nie mógł zmaleć. Czy prowadzący działania profilaktyczne powinni być zadowoleni z 10-procentowego wzrostu rozpoczynania palenia marihuany? Przy braku możliwości porównania trudno stwierdzić, czy taki wynik ma cieszyć, czy zniechęcać do realizacji programu profilaktycznego.

Wykres D przedstawia wyniki uzyskane po zastosowaniu schematu szeregów czasowych. W tym przykładzie byli badani uczniowie w piątym, szóstym i siódmym klasach przed wprowadzeniem treningu odmawiania, a następnie w ósmym i dziewiątym klasach po treningu. Nie było grupy kontrolnej ani porównawczej. Stosując ten schemat możemy stwierdzić, że po przeprowadzeniu treningu umiejętności odmawiania tempo wzrostu rozpoczynania przez uczniów palenia marihuany było wolniejsze, co jest pożądane. Schemat szeregów czasowych pokazuje, że poziom eksperymentowania z marihuaną wzrastał szybciej, zanim uczniowie zostali poddani treningowi. Taki schemat, w porównaniu ze schematem pretest-posttest przedstawionym na wykresie C, daje większą pewność, że program rozwoju umiejętności odmawiania miał wpływ na poziom używania marihuany. Jednak wątpliwości nadal pozostają. A jeśli w danej społeczności największy wzrost eksperymentowania z marihuaną przypada na okres życia

przed siódmą klasą? Jeśli ta sugestia jest słuszna, przyczyną obserwowanego stanu może być raczej naturalne dojrzewanie niż trening odmawiania. Gdy stosuje się schemat szeregów czasowych czy inny schemat quasi-eksperymentalny, ważne jest rozpatrzenie różnych możliwych wyjaśnień obserwowanych zmian i ocenienie, na ile wyjaśnienia te są prawdopodobne. Im lepiej można wykazać, że są one mało prawdopodobne, z tym większą pewnością realizatorzy programu profilaktycznego mogą stwierdzić, iż to ich działalność spowodowała wystąpienie obserwowanych wyników.

Schemat, jaki wybierzesz, w dużej mierze przesądzi o tym, na ile kompletną odpowiedź otrzymasz na swoje pytanie. Im bardziej zastosowany przez Ciebie schemat będzie zbliżony do prawdziwego schematu eksperymentalnego, tym więcej uzyskasz informacji potrzebnych Ci do sformułowania odpowiedzi na postawione pytanie.

Bardziej szczegółowe omówienie schematów znajdziesz w amerykańskiej książce Campbella i Stanley'a pt. *Experimental and Quasi-Experimental Designs for Research* (1963) albo w podobnych publikacjach (W Polsce: Sułek, 1979; Brzeziński, 1980).

Część 2. Wybierz schemat odpowiedni dla Twojego pytania i możliwości

Jak myślisz, który z omówionych schematów pozwoli Ci najlepiej odpowiedzieć na postawione przez Ciebie pytanie?

Przedstaw poniżej ten schemat w formie diagramu, określając sposób doboru grup, a także miejsce pretestu, posttestu i działań profilaktycznych.

Część 3. Określ badaną próbę

Masz już pytanie dotyczące Twojego programu i wiesz, jaki schemat pozwoli Ci najlepiej na nie odpowiedzieć. W jaki sposób dobierzesz teraz osoby do oddziaływań i badań?

Osoby badane powinny być podobne do osób z populacji, którą obejmuje Twój program profilaktyczny. Jeśli prowadzisz trening umiejętności odmawiania dla uczniów klas szóstych w trzech szkołach, powinieneś mieć w swoim badaniu dobrą reprezentację uczniów klas szóstych z tych trzech szkół.

Jeśli w jakimś dużym mieście prowadzisz warsztaty, które mają pomóc rodzicom w ustanawianiu rodzinnych zasad dotyczących substancji uzależniających, będziesz się starał włączyć do swojego badania szeroką reprezentację rodziców pochodzącą z wielu środowisk, w których odbywają się Twoje warsztaty. Taka próba pozwoli Ci dowiedzieć się, na ile rodzice, którzy uczestniczyli w Twoich warsztatach, są w stanie odpowiednio ustalić i przestrzegać w swoich rodzinach reguł dotyczących substancji uzależniających.

Jednym ze sposobów upewnienia się, że osoby badane są podobne do tych, na które oddziaływałeś, jest badanie wszystkich osób objętych Twoimi działaniami. Nazywa się to badaniem populacji. Jeśli liczba osób objętych Twoim programem profilaktycznym jest relatywnie mała — powiedzmy 90 dzieci z rodzin z problemem alkoholowym w grupach wsparcia i rozwoju umiejętności społecznych w trzech szkołach — możesz próbować objąć ewaluacją wszystkie te dzieci.

W bardzo wielu programach profilaktycznych nie można jednak pozwolić sobie na badanie wszystkich osób objętych oddziaływaniami. Jeśli prowadzi się programy profilaktyczne w szkołach we wszystkich klasach, każdego roku można mieć do czynienia z 30 000 dzieci. W takim przypadku, by oszczędzić czas i pieniądze, bada się tylko część tej populacji. Taką grupę nazywa się próbą.

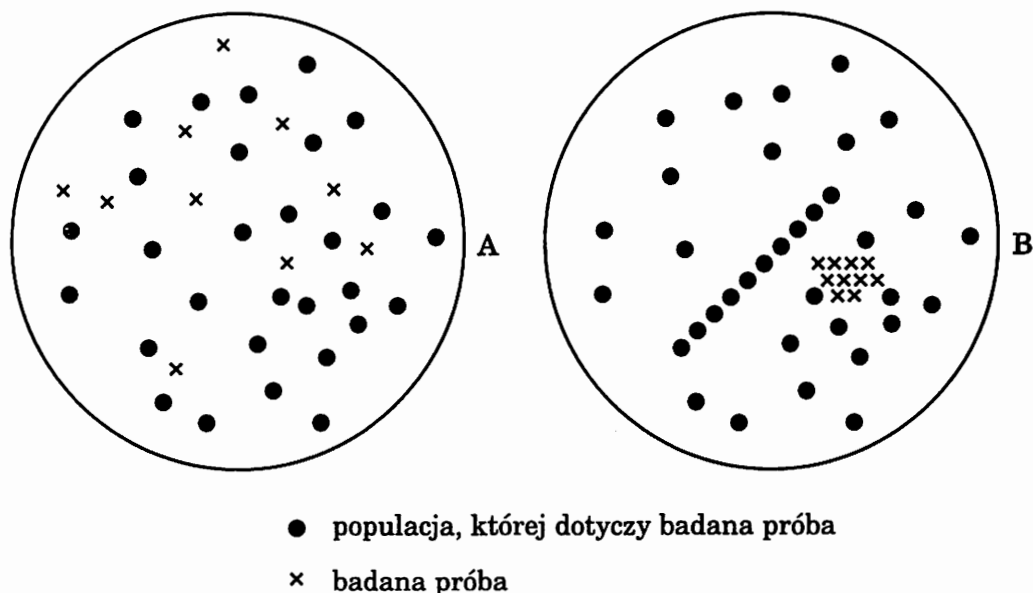
Dobór próby

Na jakiej podstawie zdecydować, czy badać całą populację czy tylko próbę? Podejmując decyzję powinieneś brać pod uwagę trzy rzeczy:

1. Aby porównać grupy po wprowadzeniu programu profilaktycznego, powinieneś mieć w każdej grupie przynajmniej 30 osób.
2. Oszacuj największą możliwą liczebność Twojej populacji. Jeśli obejmujesz swoim programem 100 albo mniej osób, to włącz je wszystkie do badań ewaluacyjnych.
3. Rozważ koszty i korzyści płynące z badania wszystkich osób objętych profilaktyką w porównaniu do badania próby, biorąc pod uwagę swój program i przyjęty schemat ewaluacji. Jeśli,

na przykład, badasz uczniów w szkołach, często może być Ci łatwiej objąć badaniami wszystkich uczniów danego poziomu klas, niż wybrać z nich tylko próbę.

Jeśli zdecydowałeś się badać tylko próbę, w jaki sposób zapewnić masz jej reprezentatywność dla całej populacji, na którą oddziałujesz? Spójrz na rycinę 6. Zarówno okrag A, jak i B pokazują próby wybrane z populacji. Jak myślisz, która z nich lepiej reprezentuje populację? Odpowiedź brzmi: okrag A. W jaki sposób wybierasz taką próbę? Poprzez dobór losowy. Jednym ze sposobów takiego doboru jest przypisywanie poszczególnym osobom przypadkowych numerów. Dobór losowy może wydawać się czymś skomplikowanym, ale w rzeczywistości jest bardzo prosty do przeprowadzenia.



Ryc. 6. Losowy i nielosowy dobór próby badanej

Odkrywanie tajemnic doboru losowego

Uzyskanie losowej próby osób jest łatwe, praktyczne i niedrogie. Jeśli zdecydowałeś, że dla Twojego badania niezbędny jest dobór losowy, oto jak możesz go przeprowadzić:

1. Sporządź listę wszystkich osób lub jednostek organizacyjnych (szkoły, klasy itp.) objętych Twoim programem, czyli całej populacji, na którą oddziałujesz. Przypisz każdej osobie lub jednostce organizacyjnej numer.
2. Jeśli prowadzisz dobór z listy, skorzystaj z bardzo prostej tablicy liczb losowych, którą znajdziesz w Załączniku C. Zamknij oczy i wskaż palcem jakieś miejsce na tej tablicy. Jeśli numer, który wskazałeś, odpowiada numerowi przypisanemu jakiejś osobie lub jednostce z Twojej listy, to ta osoba czy jednostka zostaje włączona do Twojej próby. Od tego miejsca zacznij czytać całą listę liczb losowych z tablicy. Za każdym razem, gdy trafisz na numer odpowiadający numerowi z Twojej listy, przypisz tę osobę do swojej próby. Kontynuuj tę czynność, aż będziesz miał próbę pożądanego wielkości. W ten sposób utworzysz losową próbę zaczynając od losowo dobranego punktu.
3. Jeśli przeprowadzasz dobór spośród osób właśnie rozpoczynających program, a nie z listy, włącz do próby każdą „n-tą” osobę (co 5., 10. osobę, jakkolwiek liczbę wybierzesz) aż do uzyskania próby pożądanego wielkości. Aby, na przykład, uzyskać próbę 50-osobową z populacji 500-osobowej, wybierz co 10. osobę.

Podjęcie decyzji o wielkości próby

Określenie odpowiedniej wielkości próby może okazać się skomplikowanym zadaniem. Jak wspomniano wcześniej, z grubsza rzecz biorąc, powinieneś mieć co najmniej 30 osób zarówno w grupie porównawczej, jak i eksperymentalnej. Ale nie jest to magiczna liczba. Im większa jest próba wykorzystywana do porównań, tym mniejsze prawdopodobieństwo, że wyniki, jakie uzyskałeś, są dziełem przypadku, i tym łatwiej wykazać, że różnice wywołane Twoim programem są istotne statystycznie (co po prostu oznacza, iż jest niewielka szansa, że są one przypadkowe).

Tabela wielkości prób w Załączniku C pokazuje, jakiej wielkości powinny być dwie próby, aby można było uzyskać istotne wyniki przy porównywaniu proporcji między grupami. Na przykład, jeśli badasz deklarowany przez uczniów miesięczny poziom picia alkoholu po przeprowadzonych zajęciach rozwijających umiejętność odmawiania i stwierdzasz, że 70% Twojej grupy profilaktycznej i 40% grupy kontrolnej w ogóle nie piło alkoholu, to powinieneś mieć przynajmniej 45 osób w każdej grupie, aby zyskać pewność, z prawdopodobieństwem 5 przypadków na 100, że 30-procentowa różnica między grupami nie była dziełem przypadku. A więc w 95% (95 przypadków na 100) możesz być pewien, że stwierdzona przez Ciebie 30-procentowa różnica między próbami, odzwierciedla rzeczywiste różnice występu-

jące między grupami profilaktyczną i kontrolną, z których dobierałeś próbę. (To jest właśnie ów słynny 95% poziom ufności!).

Utrata części próby (wykruszenie się)

Może się zdarzyć, że wybrałeś jakąś liczbę osób do badania, a potem okazało się, iż nie jesteś w stanie uzyskać lub zebrać wszystkich informacji od części z tych osób. Ktoś może zmienić miejsce zamieszkania, ktoś inny może odmówić odpowiedzi na pytania w Twoim kwestionariuszu albo akurat być nieobecny w szkole czy w pracy, wówczas gdy przeprowadzane są badania. Może to mieć znaczenie dla Twoich wyników. Badania pokazują, że poziom używania substancji uzależniających jest wyższy wśród osób, które często zmieniają miejsce zamieszkania, niż wśród tych, którzy nie przeprowadzają się. Jest również bardziej prawdopodobne, że wśród uczniów, którzy są nieobecni w szkole dowolnego dnia, jest więcej osób używających substancji uzależniających niż wśród pozostałych uczniów. Aby się upewnić, że na Twoje wyniki nie będzie miało wpływu wykruszenie się grupy, możesz zrobić dwie rzeczy. Po pierwsze, do badanej próby wybierz od 15% do 20% osób więcej, niż będziesz ostatecznie potrzebował. Jeśli ktoś się wykruszy, próba i tak będzie wystarczająco liczna. Po drugie, postaraj się, aby nie tracić osób wybranych do badania.

PLAN ZEBRANIA NR 2 ZESPOŁU/PERSONELU SCHEMAT EWALUACJI

UCZESTNICY: Dla małego programu — cały zespół. Dla większego programu — przedstawiciele różnych grup zainteresowanych ewaluacją, osoby z jakimś doświadczeniem lub mające znaczenie dla powodzenia ewaluacji. Jeśli korzystałeś z czyichś pomysłów dotyczących ewaluacji, np. studenta zaangażowanego przy realizacji programu badawczego, możesz zaprosić go do współpracy przy szukaniu odpowiedzi na pytania dotyczące schematu ewaluacyjnego.

CZAS: 2 godziny.

CEL: Wybranie schematu ewaluacji poprzez rozważenie każdego ze schematów (schematu pre-post, eksperymentalnego, quasi-eksperymentalnego i schematu tylko z posttestem) w kontekście pytań i środków odnoszących się do Twojego programu.

- I. **Przedyskutuj możliwości utworzenia grupy kontrolnej lub porównawczej.** (Powołaj się na rozdział 4. Dobrze by było, gdyby osoby mające kluczowe znaczenie dla ewaluacji przeczytały ten rozdział przed spotkaniem).
- II. **Rozważ schematy z grupą kontrolną, porównawczą, z pretestem, posttestem i szeregów czasowych.**
- III. **Wybierz odpowiedni schemat.**
- IV. **Poproś zespół o przeczytanie rozdziału 5 przed następnym spotkaniem.**