

Przetwarzanie danych w badaniu klinicznym

Daniel Rabczenko

d.rabczenko@iestat.pl



Instytut Edukacji I.E.



Narodowy Instytut Zdrowia Publicznego –
Państwowy Zakład Higieny

Plan

- ▶ Wstęp
- ▶ Baza danych – efekt pracy Clinical Data Managera
- ▶ Planowanie przetwarzania danych w badaniu klinicznym
 - ▶ Rozmiar próby (->plan analizy statystycznej)
 - ▶ Sposób zbierania danych -> przygotowanie CRF
- ▶ CRF – papierowy czy elektroniczny?
 - ▶ Przygotowanie, formatowanie, druk
 - ▶ Procedury zapewniające jakość danych
 - ▶ Wprowadzanie danych
 - ▶ Zapytania
- ▶ Dane laboratoryjne i inne dane zewnętrzne
 - ▶ Typy danych zewnętrznych
 - ▶ Wprowadzanie danych zewnętrznych do bazy danych
 - ▶ Standardy przekazywania danych
 - ▶ Identyfikacja
 - ▶ Problemy związane z użyciem danych z kilku źródeł
 - Normy laboratoryjne
- ▶ Dane dotyczące działań niepożądanych
 - ▶ Zgłaszanie
 - ▶ SAE Reconciliation
 - ▶ Kodowanie
- ▶ Zamknięcie bazy danych
- ▶ Początek na końcu: plan przetwarzania danych
- ▶ Współpraca osób o różnej specjalności przy przetwarzaniu danych w badaniu klinicznym

Wstęp 1

Badanie naukowe



Baza danych - 4

- ▶ **Ustalenia początkowe i wybór narzędzia**
 - ▶ Zależy od planu badania
 - ▶ Liczba ośrodków, liczba pacjentów
 - ▶ CRF papierowy, dane wpisywane centralnie
 - ▶ Dane wpisywane w ośrodkach do aplikacji offline
 - ▶ Dane wpisywane w ośrodkach do aplikacji online
 - ▶ Dane wpisywane przez ankietera w trakcie wywiadu
- ▶ **Balans pomiędzy**
 - ▶ Funkcjonalnością
 - ▶ Kosztami
 - ▶ Szybkością dostępu do wyników
 - ▶ Możliwością pomyłek
 - ▶ ...



Baza danych - 5

- ▶ **Narzędzia**
 - ▶ MS Office
 - ▶ Excel
 - ▶ Access
 - ▶ SPSS Data Entry, SAS
 - ▶ EpiData
 - ▶ Visual Basic, Python, C++...



Przykłady



Dokumentacja bazy danych

- ▶ Oznaczenie projektu
- ▶ Specyfikacja środowiska w którym utworzono bazę danych
- ▶ Identyfikatory (np.: pacjenta, wizyty, inne)
- ▶ Opis struktury
 - ▶ Szczególnie ważne, jeżeli struktura bazy danych złożona jest z więcej niż jednej tabeli
- ▶ Opis zastosowanych kodów
 - ▶ „Słownik”
- ▶ Listę sprawdzeń logicznych (jeżeli dotyczy)
 - ▶ W czasie wprowadzania danych
 - ▶ Po wprowadzeniu danych
- ▶ Źródło plików zewnętrznych
- ▶ Listę dołączonych plików zewnętrznych (jeżeli dotyczy)
- ▶ Opis łączenia plików



Testowanie bazy danych

- ▶ **Odbywać powinno się według ustalonej wcześniej procedury**
- ▶ **Sprawdzeniu podlegać powinna:**
 - ▶ Zgodność danych wpisanych z danymi z dokumentów źródłowych
 - ▶ Poprawność działania reguł sprawdzających działających w czasie wprowadzania i po wprowadzeniu danych
 - ▶ Sprawność działania bazy danych
- ▶ **Wybór danych testowych**
 - ▶ Określona w planie testowania
 - ▶ Dane odpowiednio dobrane



Planowanie przetwarzania danych w badaniu

- ▶ **Hipotezy badawcze**



- ▶ Układ badania
- ▶ Punkty końcowe
- ▶ Metody statystyczne
 - ▶ Liczebność próby



Punkty końcowe badania

- ▶ **Punkty końcowe badania**
 - ▶ Ich ustalenie jest kluczową sprawą dla sformalizowania postawionej hipotezy badawczej
- ▶ **Opisujące wynik leczenia**
 - ▶ Sukces/porażka (remisja/brak remisji)
 - ▶ Kategoryczne (CR/PR/MR/SD/PD)
 - ▶ Mierzone w sposób ciągły
 - ▶ Jakość życia
 - ▶ Wynik na końcu badania / zmiana natężenia objawów
- ▶ **Opisujące czas utrzymania się efektu leczenia**
 - ▶ Przeżycie (*survival*)
 - ▶ Czas do wznowy (*progression/disease free survival*)



Statystyka

- Wybór testu statystycznego, w sytuacji, kiedy szukamy różnic pomiędzy grupami badawczymi, przy różnym sposobie oceny efektywności leczenia, w różnych układach badawczych

Skala	Warunki dodatkowe	Liczba grup			
		Dwie grupy	Dwie grupy	Więcej niż dwie grupy	Więcej niż dwie grupy
		Zmienne nie powiązane	Zmienne powiązane	Zmienne nie powiązane	Zmienne powiązane
Interwałowa	Normalność rozkładu	Test t-Studenta (nie powiązane)	Test t-Studenta (powiązane)	Analiza wariancji	Analiza wariancji
Interwałowa	Brak normalności rozkładu	Test Manna-Whitneya	Test Wilcozona	Test Kruskala – Wallisa	Test Fridemana
Porządkowa	-	Test Manna-Whitneya	Test Wilcozona	Test Kruskala – Wallisa	Test Fridemana
Nominalna	-	Test χ^2 lub test Fishera	Test znaków lub test McNemary	Test χ^2 lub test Fishera	Test Q- Cochrana

Analiza przeżycia

- Analiza w sytuacji, kiedy przedmiotem zainteresowania jest czas utrzymania się efektu leczenia
 - Analiza czasu do wystąpienia pewnego zdarzenia
 - Zgon
 - Wznowa
 - Ale również – czas, po którym nastąpiła poprawa
- Na dane o wyniku leczenia składają się:
 - Informacja, czy zdarzenie będące przedmiotem zainteresowania (np. zgon) wystąpiło
 - Jak szybko to się stało

Liczebność próby

Informacje potrzebne do określenia liczebności próby:

- Co jest celem badania
- Jak mierzony jest punkt końcowy badania
- Jaką metodą będą analizowane dane
- Jakiego efektu spodziewamy się stosując standardową terapię
- Jaką najmniejszą różnicę chcemy wykryć i przy jakim poziomie istotności

Liczebność próby

Przykład

Badanie ma polegać na porównaniu odsetków osób wyleczonych wśród leczonych lekiem 1 i lekiem 2.

Wymagana liczba pacjentów w każdej z grup dana jest wzorem:

$$n = \frac{p_1 \times (1 - p_1) + p_2 \times (1 - p_2)}{(p_2 - p_1)^2} \times (Z_\beta + Z_{\alpha/2})^2$$

$\alpha=0.05$, $\beta=80\%$, $Z_{\alpha/2}$, Z_β – odpowiednie wartości krytyczne rozkładu normalnego

p_1 , p_2 – prawdopodobieństwa pozytywnego wyniku leczenia odpowiednio lekami 1 i 2

Statystyk „z zewnątrz” nie może nic powiedzieć na temat wielkości p_1 i p_2 !!! ➔ Konieczność współpracy badacza i statystyka



CRF

▶ Papierowy czy elektroniczny?

- | | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> ▶ Papierowy ▶ Niewielki koszt przygotowania ▶ Problemy z dystrybucją i zbieraniem ▶ Duża praca związana z wprowadzaniem danych i ich sprawdzaniem <ul style="list-style-type: none"> ▶ Jednak wykonywana przez „specjalistów” ▶ Brak problemów z IT ▶ Dane dostępne dopiero po wprowadzeniu ▶ Zdecydowanie lepszy w małych, jednośrodkowych badaniach | <ul style="list-style-type: none"> ▶ Elektroniczny (internetowy) ▶ Duży koszt przygotowania ▶ Brak problemów logistycznych ▶ Praca związana z wprowadzaniem danych przerezuona na badaczy <ul style="list-style-type: none"> ▶ Dane wpisywane przez osoby najbliższej pacjenta ▶ Możliwe problemy techniczne <ul style="list-style-type: none"> ▶ Możliwa konieczność organizacji „help desk” ▶ Wgląd w dane na bieżąco ▶ Zdecydowanie lepszy w dużych, wielośrodkowych badaniach |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



CRF

- Formularz łatwy w użyciu
- Precyzyjne sformułowanie pytań w np.
 - Czy pacjent ma chorobę serca (wymaga uściślenia)
 - Czy pacjent postępował zgodnie z instrukcjami? (zamiast: Czy pacjent nie postępował zgodnie z instrukcjami?)

Proszę porównać:

Proszę zaznaczyć grupę wieku pacjenta:
 15-24 25-34 35-44 45-54 55-64 65 i więcej

Wpisz wiek pacjenta: _____

Płeć	
Mężczyzna	<input type="checkbox"/>
Kobieta	<input type="checkbox"/>

Płeć
 Mężczyzna
 Kobieta



CRF

- Uwaga na pytania umożliwiające wielokrotny wybór

Zaznacz trzy najważniejsze uczucia związane z testowanym produktem:

Łagodzący

Kojący

Podrażniający

Obojętny

Łagodzi napięcie skóry

Likwiduje podrażnienia

Chętnie go kupię

- Uwaga na pytania otwarte

Opisz swoje odczucia związane z produktem:



CRF

- Uwaga na pytania otwarte - 2

	Choroba	Od kiedy	Czy był leczony	Leki
1				
2				
3				

	Choroba	Od kiedy	Czy był leczony	Leki
1	Włzody	12-01	tak	operacja
2	Ch.s.	2001-12	tak	aspiryna, polipiryna
3	Grypa	10-17 kw 2008	tak	leki przeciw grypowe
4	Rezygnacja	nie dotyczy	nie dotyczy	nie dotyczy



CRF – wprowadzanie danych

- ▶ Wprowadzanie danych

▶ Przepisywanie danych z dokumentacji źródłowej do eCRF lub z papierowego CRF do bazy danych

- ▶ Procedury i postępowanie związane z zapewnieniem jakości danych różne w zależności od tego czy stosowany jest CRF papierowy czy elektroniczny

▶ Wprowadzanie danych „centralne” przez przeszkolony personel

▶ Zazwyczaj dwukrotnie

- Równoczesne wpisanie danych do dwóch baz i wykonanie porównania
- „Nadpisywanie danych” do tej samej bazy



CRF – zapytania

- ▶ W trakcie wprowadzania danych/pracy związanej z „czyszczeniem” zbioru danych pojawia się często potrzeba „sięgnięcia” do danych źródłowych/kontaktu z badaczem (*query*)
- ▶ Istotna jest potrzeba opracowania systemu obiegu zapytań
 - ▶ Ustalenie odpowiedzialności
 - ▶ Rozwiązania techniczne, np.:
 - ▶ Formularz
 - ▶ Funkcjonalność eCRF
 - ▶ Ustalenie ram czasowych

▶

CRF

- ▶ Papierowy czy elektroniczny?
- ▶ Papierowy
 - ▶ Niewielki koszt przygotowania
 - ▶ Problemy z dystrybucją i zbieraniem
 - ▶ Duża praca związana z wprowadzaniem danych i ich sprawdzaniem
 - ▶ Jednak wykonywana przez „specjalistów”
 - ▶ Brak problemów z IT
 - ▶ Dane dostępne dopiero po wprowadzeniu
 - ▶ Zdecydowanie lepszy w małych, jednośrodkowych badaniach
- ▶ Elektroniczny (internetowy)
 - ▶ Duży koszt przygotowania
 - ▶ Brak problemów logistycznych
 - ▶ Praca związana z wprowadzaniem danych przerzucona na badaczy
 - ▶ Dane wpisywane przez osoby najbliższej pacjenta
 - ▶ Możliwe problemy techniczne
 - ▶ Możliwa konieczność organizacji „help desk”
 - ▶ Wgląd w dane na bieżąco
 - ▶ Zdecydowanie lepszy w dużych, wielośrodkowych badaniach

▶

Dane zewnętrzne

- ▶ Dane zewnętrzne – nie wpisane bezpośrednio do CRF, które należy włączyć do bazy danych
 - ▶ Dane laboratoryjne
 - ▶ Dane z dzienniczka chorego
 - ▶ Zakodowane dane o działaniach niepożądanych
- ▶ Istotne jest opracowanie sposobu identyfikacji danych
 - ▶ Numer pacjenta w badaniu

▶

Dane laboratoryjne

- ▶ **Dane z wielu laboratoriów**
 - ▶ Różne jednostki
 - ▶ Różne zakresy referencyjne
- ▶ **Wspólne zakresy referencyjne (generalized lab norms)**
 - ▶ Uwspólnienie jednostek, w jakich wykonywane są pomiary do jednostek z układu SI
 - ▶ Przyjęcie jednego laboratorium za standard lub
 - ▶ Obliczenie wspólnych norm na podstawie zakresów z każdego laboratorium (konceptcja „phantom laboratory”)
 - ▶ Być może konieczna analiza metodologii obliczania norm dla każdego z laboratoriów
 - ▶ Transformacja wyników

▶

Działania niepożądane

- ▶ **Obowiązek raportowania**
- ▶ **Informacje o AE, SAE muszą znaleźć się również w bazie danych**
- ▶ **SAE Reconciliation**
 - ▶ Utworzenie instrukcji związanych z wprowadzeniem informacji o działaniach niepożądanych do bazy
 - ▶ Standaryzacja informacji w różnych bazach danych
 - ▶ Okresowe przeprowadzanie porównania i uwspólniania informacji
- ▶ **Kodowanie SAE**
 - ▶ Słownik MedDra
 - ▶ Wprowadzenie kodów do bazy

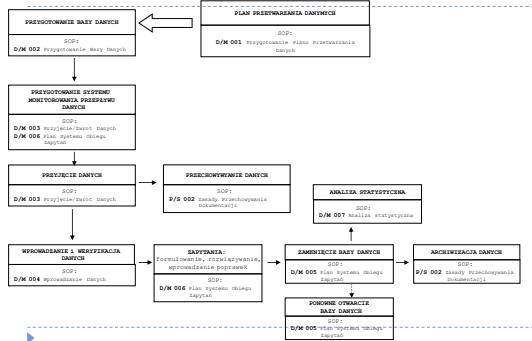
▶

Zamknięcie bazy danych

- ▶ **Konieczne jest opisanie procedury kończącej pracę nad bazą danych**
 - ▶ Wszystkie dane źródłowe zostały otrzymane i przetworzone,
 - ▶ Wszystkie wątpliwości zostały rozwiązane,
 - ▶ Zewnętrzne dane zostały przetworzone,
 - ▶ Zostały wykonane wszystkie sprawdzenia założone w planie przetwarzania danych,
 - ▶ Zostały wykonane sprawdzenia jakości danych,
 - ▶ Dokumenty i baza danych zostały zabezpieczone
- ▶ **Oraz procedury ponownego otwarcia bazy danych...**

▶

Koniec na początku – plan przetwarzania danych



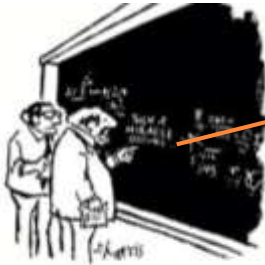
Koniec na początku – plan przetwarzania danych

- ▶ Identyfikacja czynności związanych z przetwarzaniem danych
- ▶ Identyfikacja źródeł danych
- ▶ Określenie odpowiedzialności
- ▶ Określenie ram czasowych
- ▶ Opis szczegółowych procedur

Współpraca osób o różnej specjalności przy przetwarzaniu danych w badaniu klinicznym

- ▶ Badacz
- ▶ Statystyk
- ▶ Data Manager
- ▶ Programista
- ▶ Sponsor

Dziękuję za uwagę!



"I THINK YOU SHOULD BE MORE EXPLICIT
HERE IN STEP TWO."
