Swiss TPH
Swiss Tropical and Public Health Institute
Schweizerisches Tropen- und Public Health-Institut
Institut Tropical et de Santé Publique Suisse

Identifying the key variables to be collected

Tracy Glass

Conflicts of Interest

Swiss TPH

I have received funding for membership on Data Safety and Monitoring Boards and for the preparation of educational materials from:

Gilead Sciences

Novartis

Janssen-Cilag

Outline

Swiss TPH

What data should you collect?

1. Identifying information
2. Characteristics
3. Information related to your research question

    Endpoints

    Main exposures

    Potential confounders or effect modifiers

Type of data

Identifying information

Swiss TPH

Enable you to identify individuals within study

Avoid people being included in study more than once

May need to work back to correct errors in data

Needed if you will be performing data linkage

Identifying Information

Swiss TPH

Which are identifying information?

Hospital number

Co-morbidities

Age

Ethnicity

Name

Co-medication

VL at study entry

Date of birth

Sexual orientation

Study ID number

Gender

Education level

CD4 count at study entry

Hospital name

Identifying Information

Swiss TPH

Which are identifying information?

Hospital number

Co-morbidities

Age

Name

Ethnicity

Co-medication

VL at study entry

Date of birth

Sexual orientation

Study ID number

Gender

Education level

CD4 count at study entry

Hospital name

Characteristics

Swiss TPH

- Describe the population studied
- In longitudinal studies, often collected at study entry and referred to as baseline characteristics
- Characteristics often collected:
  - Socio-demographic:
    - Date of birth, Ethnicity, Gender, Level of Education
  - HIV-related:
    - Viral load & CD4 at diagnosis (or study entry or ART initiation), AIDS diagnosis, cART regimen
  - Others:
    - co-morbidities, co-medication

Study Endpoints

Swiss TPH

- They can be referred to as outcome, event of interest, disease, dependent variable ….
- A well defined study endpoint should:
  - Be defined in advance
  - Address the primary aim of the study
  - Have biological/clinical relevance
  - Be appropriate for the population included in the trial
- Well defined study endpoints (primary and secondary) are equally important for all study designs, whether RCTs or observational studies

Example – primary endpoint

Swiss TPH

"We wish to compare the efficacy of antire
in people who uses drugs (PWD) comp
users in previously ART-naive adults i
study"

Which primary endpoint would you choose?

1) **Clinical:** New AIDS-defining event, New non-AIDS defining event, Death
2) **Virological:** Achieving VL<50 copies/ml at 1 year after starting ART, time to viral suppression, time to viral rebound
3) **Immunological:** CD4>200 cells/mm^3 at 1 year after starting ART, time to CD4 increase >100 cells/mm^3
4) **Other:** on ART at 1 year, ART switches, adherence, quality of life, toxicity

Primary and secondary endpoints

Swiss TPH

All clinical trial protocols should state one (sometimes two) pre-defined primary endpoint(s)

Main conclusions should be based on the results from this endpoint

Pre-defined secondary endpoints can also provide supportive data

For event data (i.e. diagnosis of an illness or condition) it is important to record date of event as well as fact that event occurred

Main exposures

- They can be referred to as predictors of interest, factors of interest, independent variables, …
- They should ideally also be clearly defined in advance
- In an RCT, the exposure is typically the interventions you are randomizing people to, so usually there are only one or two
- In cohort studies, the exposures are the factors that you may want to evaluate whether they predict a certain endpoint. Therefore, there is more flexibility and you may have a number of exposures.

Example – main exposure

Swiss TPH

"We wish to compare the efficacy of antiretroviral therapy in people who uses drugs (PWD) compa users in previously ART-naive adults study"

Which exposure would you choose?

1) **Source of HIV infection:** suspected source of HIV infection reported at time of diagnosis
2) **Current drug use:** any current drug use at the time of starting ART, amount of drug use
3) **Ever drug use:** ever used drugs in the past
4) **Drug program:** taking part in methadone drug program at the time of starting ART

Confounders

Swiss TPH

- Confounding is particularly an issue in observational studies, as randomization limits confounding in RCTs
- It occurs when a factor exists that is associated with **both** the **exposure** and **outcome** of interest
- Although one can never be certain that all have been accounted for, it is important to collect information on any known confounders
- It is possible to adjust for potential confounders using statistical (multivariable) models

Example - Confounders

"We wish to compare the efficacy of antiretroviral therapy in people who uses drugs (PWD) compared to non-drug users in previously ART-naive adults in an observational study"

Confounder?

Exposure:

Self-reported injecting drug use at ART start

Primary Endpoint:

VL<50 copies/ml at 1 year after starting cART

Effect modifiers

- An **effect modifier** is a variable that differentially (positively or negatively) modifies the observed effect of an exposure on the endpoint
- An effect modifier is a type of interaction
- Effect modification is a phenomenon in which the exposure has a different impact in different circumstances

Example – effect modifiers

- Monoamine oxidase inhibitors (MAOI) are used to treat depression
- People who eat certain foods, such as cheese, are at higher risk of stroke if they take MAOI
- MAOI is an effect modifier
- MAOI is NOT associated with stroke, and so is NOT a confounder

Circumstance	Exposure		Endpoint
Taking MAOI	Cheese	➡	Stroke
No MAOI	Cheese	→	Stroke

Exposures, confounders, and effect modifiers

Swiss TPH

- As measurements may change over the study period (even the exposure in an observational study!), a patient's status should be re-assessed at regular times during the study
- The frequency at which each measurement is assessed will depend on the likelihood of it changing over time, as well as the reliability of the data sources
- Example: drug use, dietary factors, smoking status, alcohol consumption

Outline

Swiss TPH

What data should you collect?

1. Identifying information
2. Characteristics
3. Information related to your research question

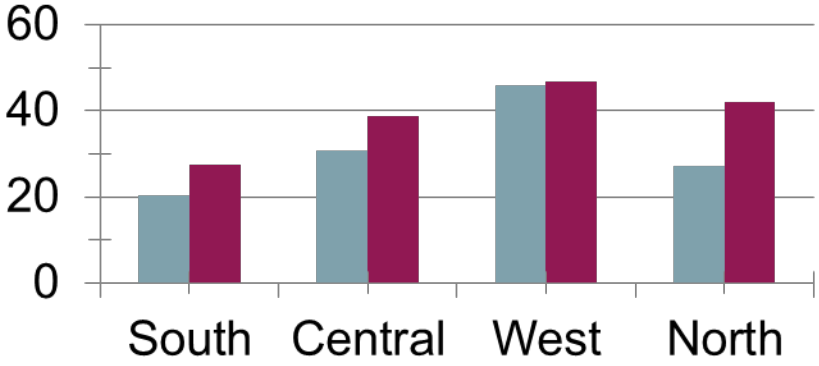
    - Endpoints
    - Main exposures
    - Potential confounders or effect modifiers

Type of data

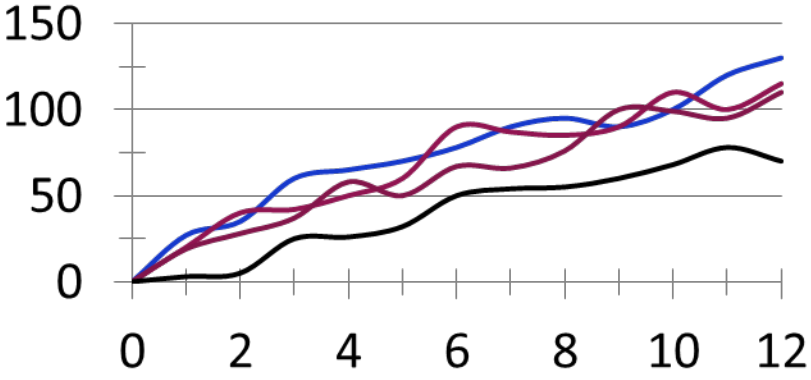
Types of data

There are two main types of data

- Categorical/qualitative



- Numerical/quantitative



Categorical data

- **Binary data**

  Two categories (yes/no, dead/alive, male/female)

- **Nominal data**

  More than two categories, no ordering to the groups (e.g. HIV exposure category, country of birth)

- **Ordinal data**

  More than two categories, some inherent ordering (e.g. CDC stage, education, some quality of life scores)

Numerical data

Swiss TPH

- **Discrete data**

- Can only take whole numbers within a given range (e.g. number of sexual partners)

- **Continuous data**

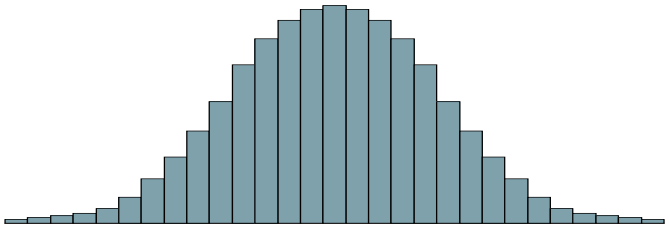
- Can take any value in a range (e.g. height, CD4 count, total cholesterol).
- Can be censored they- can only be measured within a certain range. Time to event data can only assume positive values (e.g. survival from HIV diagnosis until end of study)
- Proportions (can only assume values between 0 and 1)

The distributions of numerical data

The choice of summary statistics and the most appropriate analytical method will depend on the shape of the distribution

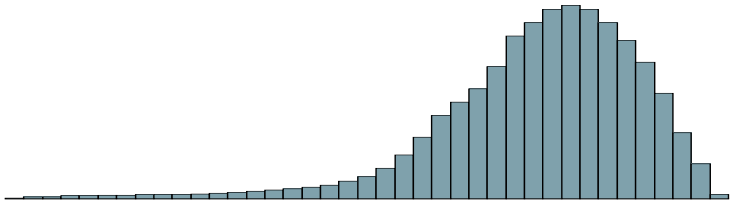
Symmetrical, bell-shaped

'Normal' distribution



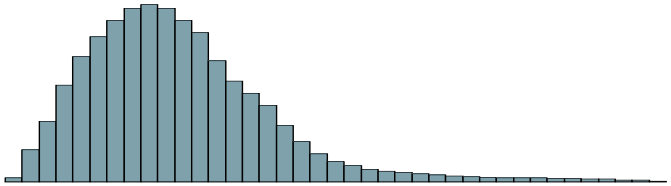
Negative skewness

eg. RNA in HIV+ not on ART



Positive skewness

common in laboratory data
eg. CD4 counts in HIV+



Uniform distribution

Equal probability of taking any value in the range

Summarizing data

Swiss TPH

- We usually quote two measures:
  - A measure of the *average* value
  - A measure of how *variable* the data are

Type of data	Average	Variability
Numerical, normally distributed	Mean	SD/variance
Numerical, skewed	Median	Range/IQR
Categorical, nominal	Mode	No suitable measure – give % in each category
Categorical, ordinal, only a few categories	Mode	
Categorical, ordinal, reasonable number of categories	Median	

Summary

Swiss TPH

It is important to consider study design and the research question to be addressed *before* beginning data collection

A clear definition of exposure, endpoint and identification of potential confounders and effect modifiers prior to the start of the study means that information on these can be collected and adjusted for