Mini Lecture 4:

What data should you collect?

What data should you collect?

Data needed to perform research study will be specific to the research question but can be broken down into

- Identifying information
- Baseline characteristics
- Information related to your research question
- Safety, tolerability, side-effects

Personal data

“ data which relate to a living individual who can be identified

(a) from those data, or

(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller ”

Data Protection Act, 1998

Personal data

Examples of personal data

Personal data

Examples of personal data

- Name
- NHS Number
- Clinic/hospital number
- Address
- Postal/zip code
- Date of birth

Sensitive data

Data consisting of information relating to

- *racial or ethnic origin*
- *political opinions*
- *religious beliefs*
- *membership of a trade union*
- *physical or mental health*
- *sexual life*
- *criminal offences*

Sensitive data

Examples of sensitive data

- Ethnicity
- Religion
- Sexuality
- HIV-positive status
- Salary
- Bank account number

Personal & sensitive data

- Collecting personal and sensitive data will require informed consent and ethical approval
- Want to collect as much personal/sensitive information as you need to conduct your study but not collect and hold unnecessary data

e.g. location

Address → postcode → region → urban/rural → ??

What data should you collect?

- Identifying information
- Baseline characteristics
- Related to your research question
- Safety, tolerability, side-effects

Identifying Information

- 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

Examples

Study ID number, name, date of birth, hospital

Example

In a cohort study you wish to address the following research question:

Q: Do people who start ART with a CD4 count >500 cells/mm^3 experience a higher rate of ART toxicity than people with a CD4 count ≤500 cells/mm^3 ?

What *identifying information* would you collect?

What data should you collect?

- Identifying information
- Baseline characteristics
- Related to your research question
- Safety, tolerability, side-effects

Baseline characteristics

- Describe characteristics of population studied
  - Generalisability
  - Check groups are balanced (RCT)
- Capture information on important factors, potential confounders and sub-groups of interest

Examples

Demographics: Age, sex, ethnicity, socio-economic

HIV-related: CD4 count, viral load, ART, AIDS

Other: Co-morbidities, co-medication

Example

In a cohort study you wish to address the following research question:

Q: Do people who start ART with a CD4 count >500 cells/mm^3 experience a higher rate of ART toxicity than people with a CD4 count ≤500 cells/mm^3?

What *baseline characteristics* would you collect?

What data should you collect?

- Identifying information
- Baseline characteristics
- Related to your research question
- Safety, tolerability, side-effects

Data related to research question

These are any data relating to either your

1. Exposure
2. Outcome
3. Confounding factors

which may be measured at baseline and/or during study follow-up

Data related to research question

- 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
- Retain detail – continuous variables can be grouped into categories but can’t work backwards

Example

In a cohort study you wish to address the following research question:

Q: Do people who start ART with a CD4 count >500 cells/mm^3 experience a higher rate of ART toxicity than people with a CD4 count ≤500 cells/mm^3?

What *data related to your research question* would you collect?

Example

What *data related to your research question* would you collect?

- Exposure:
- Outcome:
- Confounding factors:

Example

What *data related to your research question* would you collect?

- Exposure: Date of starting ART, CD4 count at start of ART
- Outcome:
- Confounding factors:

Example

What *data related to your research question* would you collect?

- Exposure: Date of starting ART, CD4 count at start of ART
- Outcome: ART-related toxicity
- Confounding factors:

Example

What *data related to your research question* would you collect?

- Exposure: Date of starting ART, CD4 count at start of ART
- Outcome: ART-related toxicity
- Confounding factors: Reasons for starting ART, comorbidities, ethnicity, pregnancy

What data should you collect?

- Identifying information
- Demographic information
- Related to your research question
- Safety, tolerability, side-effects

Safety and tolerability

- Mostly applicable in a trial setting where you are testing an intervention
- May be a secondary outcome of study
- Usually measured repeatedly over study follow-up

Examples

Laboratory tests, side-effects, discontinuation of drugs

Safety and tolerability

In a **randomised controlled trial** you wish to address the following research question:

Q: Do people who start ART with a CD4 count >500 cells/mm^3 experience a higher rate of ART toxicity than people with a CD4 count ≤500 cells/mm^3?

What *safety and tolerability data* would you collect?

Summary

- Need to carefully consider in advance what data you need to answer your research question
- Appropriate steps should be taken to protect personal and sensitive data
- Retain as much detail as possible in data collected

REALITY Study