INITIAL CONSIDERATIONS ABOUT DATA

- What, When and Who
- Codebook

Initial considerations

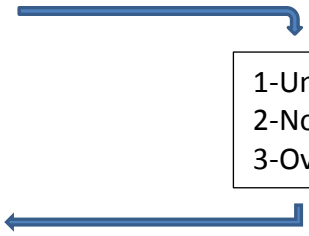
Some things to consider before setting up any type of electronic data capture platform

- What to collect
- When to collect it
- Who will be collecting
- Data representation / validation (Codebook)

Initial considerations

What to collect?

- What is required to test the hypothesis in the protocol
  - Plus obvious and less obvious confounders
- How to best capture a given data element
  - Numerical
    - Continoues – weight (kg with 1 decimal)
    - Discrete – units of alcohol pr. week
  - Categorical
    - Nominal – hair color
    - Ordinal – ‘always, sometimes, rarely’
- Use validated methods/forms when possible
  - E.g. pain, depression, disability, comorbidity
- What is the source of the data
  - Patient, patient record, lab, imaging, …



Initial considerations

When to collect?

- What are the events of the study where data is collected in some way
  - Baseline, visit 1,2,3, Questionare, …, end of study, followup
- What data is collected at those events
- Plan for additional expected or unexpected events
  - Extra visits, hospitalizations, adverse events

Initial considerations

When to collect - event grid

Data Collection Instrument	Inklusionssamtale (1)	Telefonsamtale 1 uge (2)	Konsultation 1 mnd (3)	Sgp. besøg 1 (4)	Sgp. besøg 2 (5)	Sgp. besøg 3 (6)	Sgp. besøg 4 (7)	Sgp. besøg 5 (8)	Sgp. besøg 6 (9)	Sgp. besøg 7 (10)	Afsluttende konsultation (11)	Løbende (12)
Inklusion (X)	✓											
Baseline	✓											
Sf36 Version 2	✓						✓				✓	
Cldq Spoergeskema	✓			✓		✓		✓			✓	
Tlf Kons Uge 1		✓									✓	
Konsultation Efter 1 Maaned			✓									
Sygeplejerske Konsultation				✓	✓	✓	✓	✓	✓	✓		
Afsluttende Konsultation											✓	
Hndelsesskema (X)												✓
Dataark												✓
Medicinskema (X)												✓
Ekstra Besg												✓
End Of Study												✓

Initial considerations

Who collects?

- Who will be collecting/entering data?
  - Nurses, doctors, patients themselves via questionares, diaries
- Which sites are involved?
- Which level of access is required?
  - Read / Write/ Create / Lock / Export /…
  - Restrictions on which patients can be accessed in multisite studies

Initial considerations

Other things

- Randomization integrated in database
  - Type and details (strata, blocksizes,..)
- Double data entry of data on paper
- Could some of the data be imported from existing digital format

Create a Codebook

For every data element to be collected consider the following points:

- Variable type (categorical, numerical, …)
- Variable name (consistent naming convention)
- Label (concise and targeted towards intended audience)
- Field type (text, radiobutton, checkbox, date,..)
- Entry format validation (dates as Y-M-D or D-M-Y, 1 decimal, 2 decimal, email)
- Value range validation (height: 140-220 cm)
- Precision for numeric data (integer, 1 decimal, ..)
- Units of measurement (kg, mmol, ..)
- Available choices and codes for those choices
- Required / essential data points
- Missing data (how to handle it, if it is relevant)
- Is the data element sensitive (cpr-number, name, email,..)
- Branching logic (logic for skipping the data point based on other answers)
- At which events in the protokol will this datapoint be captured
- Which logical schema/instrument does this data element best belong to (inclusion, visits, lab, ..)

Create a Codebook

In a spreadsheat, layout all your data elements as rows, with columns representing relevant information about each variable. Below a simple example

Variable name	Variable Label	Variables types	Legal values	Format	Value labels	Missing
First_name	First name	Text				
Last_name	Last name	Text				
Sex	Sex	Categorical / discrete	1,2		1, male 2, female	
Bl_date	Date baseline visit	Date	01-01-2015 – 31-12-2016	D-M-Y		
Bl_age	Baseline age	Numerical / discrete	18-99	Integer		
Cur_smoke	Is patient currently smoking	Categorical / discrete	1,0,55		1,yes 0,no 55, not disclose	999
Weight	Weight (kg)	Numerical / continues	40.0 – 200.0	1-decimal		999