Lessons learned from a Cluster randomised trial using multisource hospital routine data

Maria Elstad

Content

- Background and trial design
- Lesson 1: Plan the data collection
- Lesson 2: Check the data
- Lesson 3: Same source ≠ same format
- Lesson 4: There is a light at the end of the tunnel

The DESIGN trial

Detection of Small for Gestational age Neonate

Background

- High rates of stillbirth in UK, compared to other developed countries
- Association between small-for-gestational-age (SGA) foetuses and stillbirth
- Improved antenatal detection of SGA may reduce stillbirth

Trial design

- Cluster randomised trial
  - 13 maternity units
    - 7 assigned to implementation of the Growth Assessment Protocol (GAP)
    - 6 assigned to standard care
- Aim: to evaluate the effect of the GAP programme
- Primary outcome: antenatal ultrasound detection of small for gestational age (SGA) infants, found to be SGA at birth
- Timeline:

Baseline period (1 year pre-trial)	Implantation of GAP (approx. 1 ½ years)	Trial period (6 months)
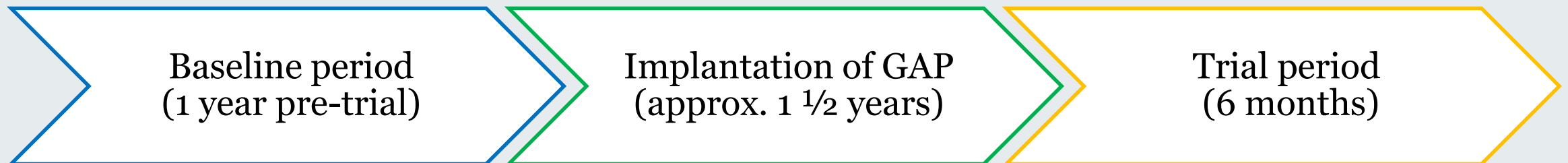
Lesson 1 : Plan the data collection

What information do we need?

Data collection

- Pre-specified variable lists
- Access to electronic hospital records
  - Local hospital teams extracting the data
  - Site visits
    - Identifiable information
      - Calculate age
      - Calculate IMD/LSOA/MSOA score
      - Study ID
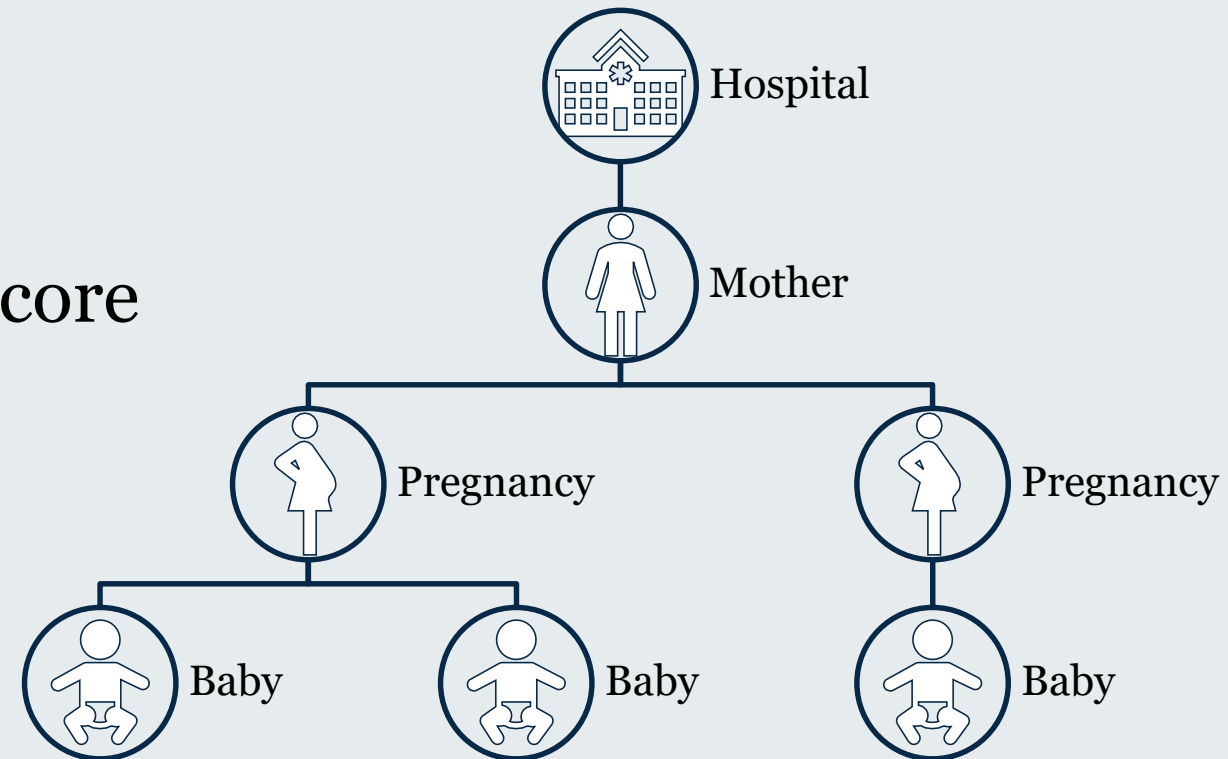
Pseudo-anonymisation tool

- Two key objectives
  - Standard software available on most NHS computers
  - Same ID for the same person regardless of hospital
- Choosing reliable identifiable variables
- Random component
- Validation through simulations
- Implementation
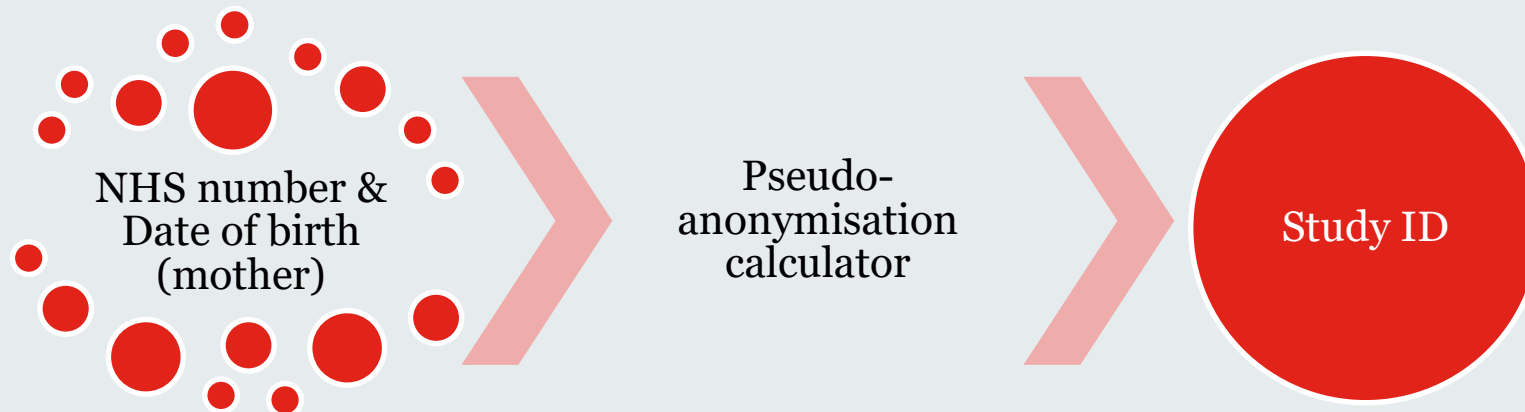
NHS number & Date of birth (mother) → Pseudo-anonymisation calculator → Study ID

Data sources

Data Request

	Maternity Information System	Ultrasound system	Neonatal intensive care system	Hospital activity system
Demography (mother)	x	x		
Primary outcome data	x		x	
Maternal data	x			x
Neonatal Clinical	x		x	
Health economics	x		x	x
Ultrasound data	x	x		

Lesson 2: Check the data

If possible, plan to download the data at least twice

Example extract

- One hospital
- Maternity data system
- Demography
- Baseline time period

Maternity overview

- Missing NHS number: 32
- More than one pregnancy during the study period: 5
- Multiples:
  - Twins: 141
  - Triplets: 7
- True duplicates: 0
- Total number of unique mothers: 6418
- Total number of babies: 6578

Demography data requested

Variables requested	
Age	Hypertension
Smoking	Systemic Lupus Erythematosus
Education	Antiphospholipid syndrome
Ethnicity	Pre pregnancy diabetes
Country of birth	Previous obstetric history
Parity	Previous gestational diabetes
Maternal height	Previous large for gestational age infant
Maternal weight	Previous small for gestatinal age infant
BMI at booking	Current pregnancy risk factors
Previous medical history	Antenatal pbr tariff

Demography data received

Variables received	
Age	Hypertension
Smoking	Systemic Lupus Erythematosus
Education	Antiphospholipid syndrome
Ethnicity	Pre pregnancy diabetes
Country of birth	Previous obstetric history
Parity	Previous gestational diabetes
Maternal height	Previous large for gestational age infant
Maternal weight	Previous small for gestatinal age infant
BMI at booking	Current pregnancy risk factors
Previous medical history	Antenatal pbr tariff

Matching received data to the request form

Different names

- Age -> Age at Delivery
- Smoking -> Smoker at Booking?
- Parity -> ParaBeforeDelivery
- Previous medical history -> MedicalDiagnosis codes
- Systemic Lupus Erythematosus -> Autoimmune

Data completeness

Variable	No.	Completeness*
Age at Delivery	6418	100%
ParaBeforeDelivery	6418	100%
Current Pregnancy Risk Factor	6398	100%
Diabetes	5600	87%
Smoker at Booking?	5560	87%
Height at Booking Observation	5532	86%
Weight at Booking Observation	5492	86%
Hypertension	5436	85%
BMI at Booking	5410	84%
Country of Birth	4903	76%
Autoimmune	4754	74%
Mother Ethnicity	4310	67%

*Based on a total of 6418 unique mothers

Common data challenges we encountered

- Duplicate records
- Different units e.g. height in metre, cm or ft
- Variation is categorical cut-offs
  - E.g. BMI, Ethnicity classifications
- Missing variables
- Missing data

Lesson 3: Same source ≠ same format

The software differs from hospital to hospital. You might need harmonization strategy

Software systems used at the sites

Maternity information system	Ultrasound information system	Neonatal information system	Hospital administrative system (appointments / admissions)
Medway Maternity K2 E3 Cerner Euroking CMIS EPR Badgernet Maternity (Clevermed)	Astraia Viewpoint (GE Healthcare) CRIS Solitorn RIS / PACS	Badgernet Neonatal (CleverMed)	Medway PAS CMIS CareCast EPR APAS OASIS iClip

Harmonisation process

Prefix for variable origin,

m=maternity
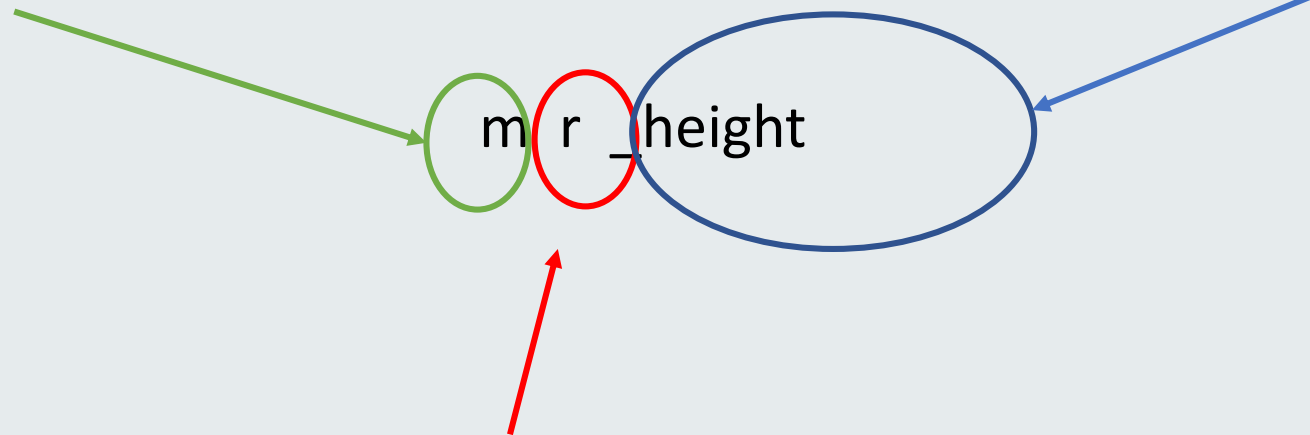
Variable descriptor

m r _height

Prefix for variable type, r=raw

Harmonisation process

Codebook -created and updated by the clinician

Variable name	Title	Variable Explanation	Variable option	Hospital X
mc_studyid	Study ID	ID of the study participant		
mc_bmicat	BMI category	Categorised from mc_bmi	1 = <18.5 2 = 18.5-24.99 3 = 25.0-29.99 4 = 30.0-34.99 5 = 35.0-39.99 6 = 40.0+	
mc_timeperiod	Birth within baseline, transitional or comparison period		0 = baseline 1 = trial 2 = comparison	Baseline = 3/11/15 - 2/11/16 Trial = 3/11/16 -31/8/18 Comparison = 1/9/18 - 28/2/19"
mr_dm	Pre-existing diabetes	Maternal existing Type 1 or Type 2 diabetes mellitus	As provided	motherendocrinemetabolicmedic Diabetes Hyperthyroidism Hypothyroidism None Other

Harmonisation process

13 hospitals *305 variables (from 4 sources)

*X versions ⟳

= countless hours of fun

Lesson 4: There is a light at the end of the tunnel

Advantages

- Easy to recruit
  - Opt. out policy
- Large dataset
  - Data was collected on 182,052 babies from 178,350 pregnancies in 165,397 unique women
  - Approximately 3 years of data per hospital
  - 305 variables
- Time
  - Active engagement throughout –no 'passive' time
  - Baseline data from before the study started

Thank you!