Government of **Western Australia**
Child and Adolescent Health Service

CAHS Research Education Program

Research Skills Seminar

Data Collection and Management

23 July 2021

Presented by

Associate Professor, Sue Skull
Head – Research Education Program
Deputy Director – Department of Research

CAHS Research Education Program
Research Skills Seminar Series
✉ ResearchEducationProgram@health.wa.gov.au
🌏 https://cahs.health.wa.gov.au/ResearchEducationProgram

Child and Adolescent Health Service, Department of Research

Department of Health, Government of Western Australia

Government of **Western Australia**
Child and Adolescent Health Service

Data Collection and Management

PRESENTATION SLIDES

CAHS Research Education Program
Research Skills Seminar Series
☎ (08) 6456 4585 ✉ ResearchEducationProgram@health.wa.gov.au
https://cahs.health.wa.gov.au/ResearchEducationProgram

Government of Western Australia
Child and Adolescent Health Service

Data Collection & Management

Responsibilities and practical strategies for database development, data entry, cleaning and storage

23 July 2021

Presented by Associate Professor Sue Skull
Head, CAHS Research Education Program

Research Education Program | Research Skills Seminar Series

1

Acknowledgement of country

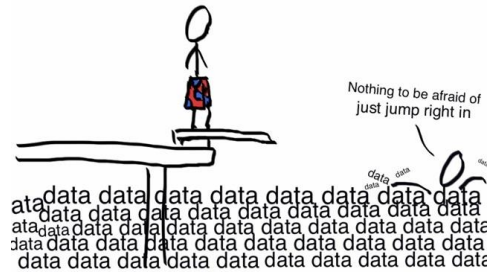
I would like to acknowledge the traditional custodians of the land, the Noongar Whadjuk people, and pay my respects to their elders, past, present and future.

2

Overview

- Good data management – why we need it
- Responsibilities – collection, storing, archiving
- Strategies for data management
- Basics of setting up a database
  - Minimising data errors
- Resources

Good Clinical Practice training



3

3

Why good data management?

- The most valuable output from research = DATA
- Bad data management: can't answer Q, inappropriate application
- Properly managed data → massive benefits
- Must be standard practice/planned from outset

4

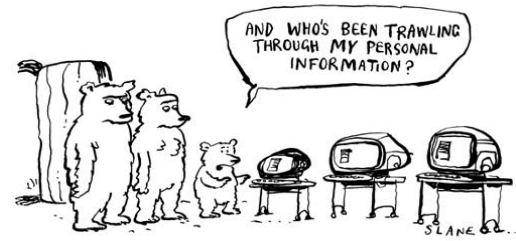
4

Why good data management?

→ High quality data

- Accurate, complete, retrievable, secure, available, identifiable

- Meaningful answer to your research question
- Prevent disasters – loss of data, confidentiality breaches
- Meet legal and institutional obligations
- Facilitate audits
- Efficiency!



5

5

Data Management and Responsibilities

6

6

Data Management Compliance

- Know relevant policies/codes
- **Good Clinical Practice**
- Consequences for failing
  - Funding bodies
  - Legal*
  - Reputation
- Increasingly required:
  - Funding bodies
  - Grant applications
  - Government
  - Institutions

Have a plan! – get input from IT, data services, biostatistician etc.
Do some Good Clinical Practice training

7

7

Guidelines and Codes

- NHMRC: Australian Code for the Responsible Conduct of Research (Jun 2018)
- National Statement on Ethical Conduct in Human Research (Jul 2018)
- Commonwealth Privacy Act (S95 and S95A) (May 2020)
- Freedom of Information Act 1992 (WA)
- Competencies for Australian Academic Clinical Triallists (May 2018)
- **WA Health Research Governance Policy and Procedures**
  https://ww2.health.wa.gov.au/Health-for/Researchers-and-educators/Research-governance

8

8

Data Management Policies

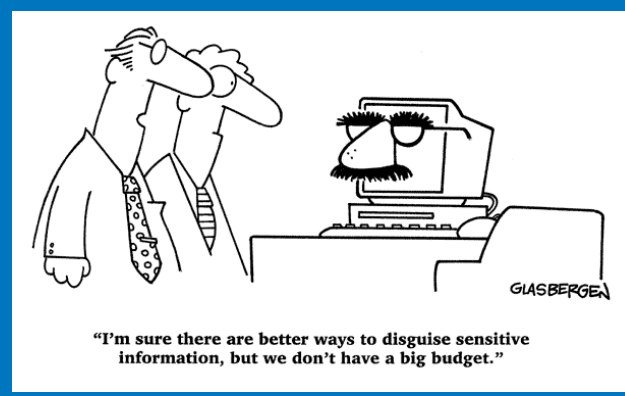
CHECK YOURS!
Follow relevant legislation and generally include:

- Ethical data collection and maintenance
- Data ownership: staff/student's institution / government
- Keeping an inventory of all research data, methodology
- Listing all variables with meaningful descriptions
- Password protection requirements
- Backup e.g. electronic databases onto institutional network
- Requirements for retention periods / access / disposal

9

9

Strategies for Good Data Management



"I'm sure there are better ways to disguise sensitive information, but we don't have a big budget."

10

10

So how to we get “good” data?

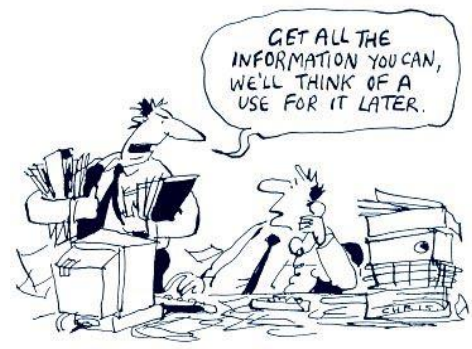
- Collect high quality data
- Data management plans
- File organisation and naming
  - Documentation of metadata
- Database design
  - Data entry, validation, cleaning
- Confidentiality, storage, backup

11

11

Quality Data Collection

- Collect only the data you need!
- Instrument design: shortest list of variables needed
- Sampling strategy
- Maximise completion rate
- Standardised operating procedures
- Piloting
- Training
- Monitoring

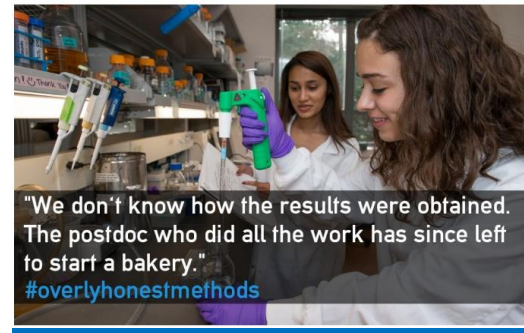


12

12

Data management plan

- Increasingly a requirement
- Document project information
- Leave out at your peril!
- Invaluable resource for
  - New project members
  - Audit
  - Post-project review



Avoid this!

13

13

Data management plan – Check List

- Data to be created

*An **Institutional** approach is best*

- Data owners and stakeholders
- File formats and organisation: names, folders, versions
- Metadata – standards, descriptions – data dictionaries
- Access and security
- Storage, backup, archiving, disposal
- Sharing and reuse
- **Responsibilities**
- Budget – hardware, curation, archiving etc.
- Protection of IP…..Other?

14

14

Good File Organisation and Naming

- Standardisation creates uniformity within/across projects
- Good Clinical Practice (GCP) outlines strategies, requirements
- Allows for easy retrieval, version control
- Maintain a single master copy of your data file
- Always make a copy before entry/edits
- Name your copies using the current date
  e.g.DataFile_2017-08-01.rec
- Define an intuitive directory structure e.g.
  - /projectA/data/masterfile
  - /projectA/data/analysis
  - /projectA/data/backups… etc.

15

Documentation of Metadata

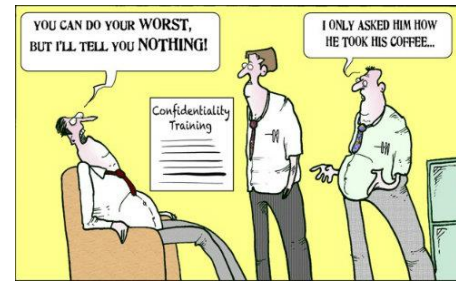
- Information that describes project data
- Ensures you remember what you did & when & why
  - Study outline
  - Functions of all files including data collection document
  - Data creation: time, author, program, location
  - Data dictionaries
  - Changes made and why – cleaning and analysis files
  - Backup processes, references to related data

16

16

Confidentiality and Security

- Individual /state institutional policies apply e.g. WA Health
- Failure to maintain confidentiality can result in prosecution
- Password protection
- Locking of hard copy data
- No removal of hard copies
- **Use of codes to de-identify data**
- Separate name and contact details
- Data transport care*
- Backup…



17

17

Back-up

- Credible written backup strategy
- Back up at the end of each day – automate if possible
- Keep back ups in more than one location
- Ideally hard drive or cloud
- Data repositories
- File format planning*

"We back up our data on sticky notes because sticky notes never crash."

18

18

Data Sharing and Reuse

- Plan from the start of a project
- Individual consent may be required
- Open access, vs institutional/ national/ international repository
- Data license - outlines access to and use of data
  e.g. creativecommons.org.au
  http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

19

19

Preservation and Archiving

- All data: electronic and paper
- All cleaning and analysis command files
- All related documentation
- Date of creation, expected destruction
- Minimum storage
  - Adults usually 5 years
  - Children 5 years after last reference or until child turns 25y
- Hard copies off site OK if accessible/infrequently needed
- **"Transfer to State Records Office in WA"***
  → approved sites
  - Retain: administrative or functional records longterm

20

20

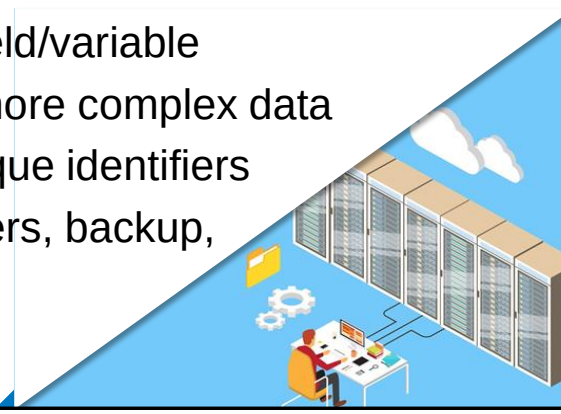
Basics of Database Set-Up

21

21

What is a Database?

- For entry, storage and manipulation of data
- Data are stored in tables
  - Each row represents a case/record
  - Each column represents a field/variable
- Relational databases great for more complex data
  - link multiple tables using unique identifiers
- Web-based enables multiple users, backup, access to subsets etc



22

Database Software *institutional level

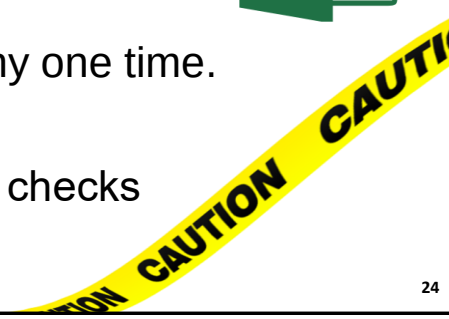
EpiData	Small, free portable program
EpiInfo	Small, free, does analysis, easy to use
REDCap*	Free, stored locally, web access, online tutorials, standard forms, reports, front end checks and balances, multiple users
Medrio	Clinical trial data entry, good security, cloud-based
Webspirit	Clinical trials – available through PTNA
SPSS Data Entry	Stand-alone product, allows validation
Qualtrics	Online tool for creating surveys, Australia-based, paid option
iApply	Online forms such as surveys
+ Oracle, MySQL, Access, lots of others	

23

23

Excel Users Beware!

- Not a database program – generally not recommended
- VERY EASY TO DESTROY YOUR DATA IRREVOCABLY
- Comprehensive data checking required
- Unable to enforce uniqueness for an identifier
- Not a relational database
- Only one person can access a file at any one time.
- Need to be careful with dates
- Very few validation rules, can't do logic checks
- No auto backup



24

24

Survey Monkey Users Beware!

- You don't own the data
- Unclear where data are
- You may be breaching institutional/government policy
- Generally NOT recommended for research data
  - Never for sensitive, identifiable/re-identifiable health data
  - Don't use, or use with extreme caution

25

25

REDCap

REDCap
Research Electronic Data Capture

- The future for data entry
- Free, intuitive, secure, collaborative, relational
- Increasingly used throughout WA
- Data entry package of choice for Dept of Health
- Resources: **CAHS Research Education Program**
  - Seminar, Workshops x4, Handouts, Access and support

26

26

Data Dictionaries

Before creating your database

- Variable names & descriptions
  - Explanation, data type, units
- Validation/coding rules
  - Code for categorical variables
  - Ranges for continuous data and dates
  - Codes for missing data
- Table names, relationships
- Relationships with other databases
- Documents database changes throughout the project
- ***Living document***

27

27

Naming Variables

Variable names must

- Be unique
- Be informative but short
- Have no spaces or punctuation marks (*underscore or CamelCasing can sometimes assist)
- Start with a letter (usually)
- Be compatible with data entry + statistical packages

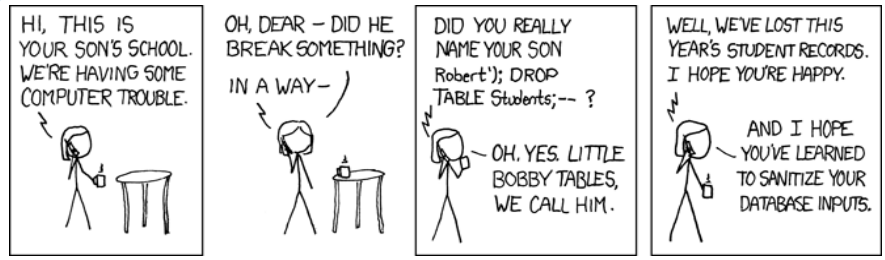
28

28

Naming Variables

Check:

- Maximum characters allowed 20+? 10? 8?
- Allowable characters _
- Whether the package is case sensitive (e.g. Stata)
- Any package-specific special names that can't be used?

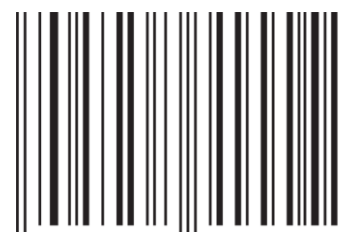


29

29

Unique Identifiers

- Each record must have a unique identifier
- Generally assigned by the researcher/package
- Refer to specific records without identifying names etc.
- Not the hospital record number
- Link data between database tables
- Generally a single field (e.g. studyid)
- REDCap always includes this field first
  - Consecutive numbers



30

30

Selecting Data Types

- **Numeric** – where calculations needed
  - continuous data
  - categorical data even if ordinal
- **Text or "string" variables**
  - Consistency with spelling, case etc. important
- **Dates** – select a date format
  - Don't use string codes
  - Don't enter D/M/YY as separate variables
- **Age/Time variables:** let the package do the work
  - Don't combine text and numbers (or scales)

31

31

Age

age	Freq.	Percent	Cum.
3 years	1	11.11	11.11
1 year	3	33.33	44.44
2 years	1	11.11	55.56
3 mnth	1	11.11	66.67
4 yrs	1	11.11	77.78
5 years	1	11.11	88.89
6 months	1	11.11	100.00
Total	9	100.00	

32

32

Case Sensitivity

gender	Freq.	Percent	Cum.
Female	2	11.76	11.76
Male	1	5.88	17.65
female	5	29.41	47.06
male	9	52.94	100.00
Total	17	100.00	

33

33

Codes

- Assist those recording or entering data
- Reduce misinterpretation and errors

Do you speak a language other than English at home?
Are you a permanent resident or citizen of Australia?
Do you consider yourself to have a disability?
Code these responses as No = 0; Yes = 1

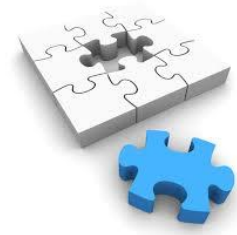
AQTF 2007 Learner Questionnaire Code Book
Commonwealth of Australia

34

34

Missing Data

- Include a code for missing data
- Collate and deal with missing codes
- **Must never be a valid response**
- **Must be changed to missing values before analysis**
- Unique ID and eligibility criteria must never be missing
- Conventionally represented by
  - “9” for one digit variables
  - “99” for two digit variables etc.
  - 99/99/9999 for dates



35

35

Data Entry Queries

- Include codes for data entry queries
  - Illegible, unclear, inconsistent, impossible
- Use a value that is not a valid field response
  - e.g. 77 or 77/77/7777
- Not for unique identifier fields
- Deal with all queries prior to analysis

More in your handout….

36

36

General Data Issues

- Consistent date and time formats
- Enter raw data
  - Keep individual repeated measurements
  - Calculate later using analysis package
- Identify sections with repeating information
- Decide on wide or long table form
  - wide – one record per subject
  - long – multiple records per subject
  - *important for analyses/package

37

37

Wide Table – Year 8 HPV Vaccine

idno	Date1 (T1)	…	Date2 (T2)	…	Date3 (T3)	…
1	6/2/03		6/4/05		7/08/03	
2	6/2/03		6/4/05		9/9/99	
3	6/2/03		19/4/03		30/08/03	
4	6/2/03		9/9/99		9/9/99	
5	6/2/03		6/4/03		7/08/03	
6	9/9/99		9/9/99		9/9/99	
⋮						

Consider if time points the same for all subjects, at similar intervals

38

38

Long Table – Year 8 HPV Vaccine

idno	dose	date	…
1	1	6/2/03	
1	2	6/5/03	
1	3	7/8/03	
2	1	6/2/05	
2	2	6/5/05	
2	3	9/9/99	
⋮			

Consider if same variables collected at each time point
Discuss with a data manager or statistician!

39

39

Dictionary Variable Definition example 1

studyid

- Each record should have a unique identifier
- Id numbers should be written on every form
- Specify the range of id numbers on the coding sheet
- There should be no missing values

Variable name	Description	Data type	Values/Rules
studyid	Participant's unique study id number	number	Must be unique 1001 – 2000

40

40

Dictionary Variable Definition example 2

Oralfeed
What was the baby fed?

☐ Breast milk 1 ☐ Formula 2 ☐ Breast milk & formula 3

- **Categorical data (nominal)**

Variable name	Description	Data type	Values/Rules
oralfeed	Oral feed type	number	1 = breast milk 2 = formula 3 = breast milk & formula 7 = query 9 = missing

41

41

Dictionary Variable Definition example 3

weight2

Weight at 2 years: ☐☐.☐ kgs

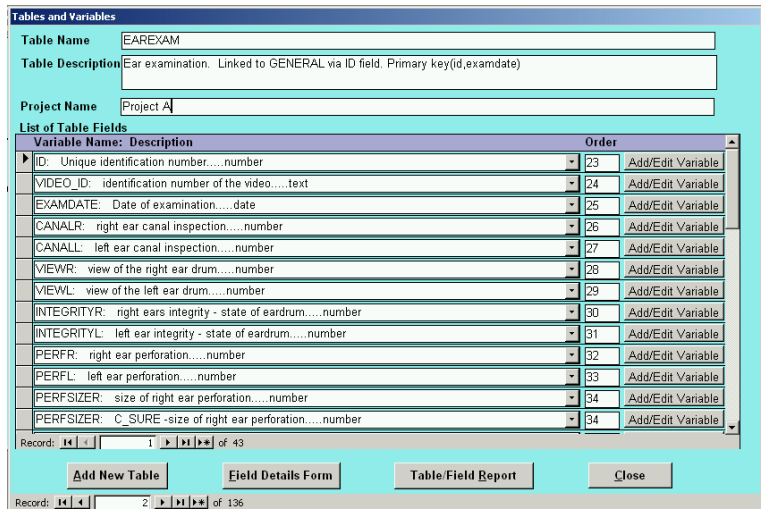
- Continuous variable
- Include unit of measure
- Specify plausible range
- Use numbers outside plausible range for query/missing

Variable name	Description	Data type	Values/Rules
weight2	Weight at 2 years (kg)	number	< 20 77 = query 99 = missing

42

42

Data Dictionary Tool – Table Example



43

43

Data Dictionary - Report Example

CLIN_OBS Table

Clinical observations starting from when the child was admitted to ISOP (7B). Primary key: (study_id, obs_time_pt, obs_date). Relationships: Links to DEMOGRAPHIC table via study_id.

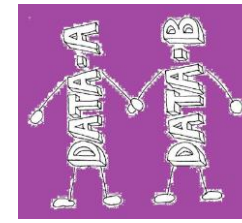
Variable name	Description	Data Type	Values/Rules
STUDY_ID	Study identification number issued to child from randomisation form. 1000's = Indigenous 26 wks & less; 2000's = non-Indig 26 wks & less; 3000's = Indig > 26wks; 4000's = non-Indig > 26 wks	number	1000-4000
OBS_TIME_PT	Clinical observations time point	number	1=Baseline hosp obs
			2=Enrolment obs
			3=12 hourly obs
OBS_DATE	date and time (24hr) of this clinical obs	date	
TEMP	Temperature (deg C)	number	25-45
PULSE	Pulse rate (beats per minute)	number	50-250
RESP	Respiratory rate (breaths per min)	number	20-120
OXY	Supplemental Oxygen (L/min)	number	0-10
RA_SAT	Oxygen saturation on room air (%)	number	60-100

44

44

Relational Databases

- Link tables using unique identifiers (primary key)
- Allow data to be easily extracted and manipulated
- Each table is like an academic paragraph
  - Deals with only one subject
  - Variables within it must relate to that subject
  - And no calculated fields
- Each field should hold just one piece of data
- Repetition of fields in a table indicates a need for another table: a one-to-many relationship



45

45

Non-Relational Database

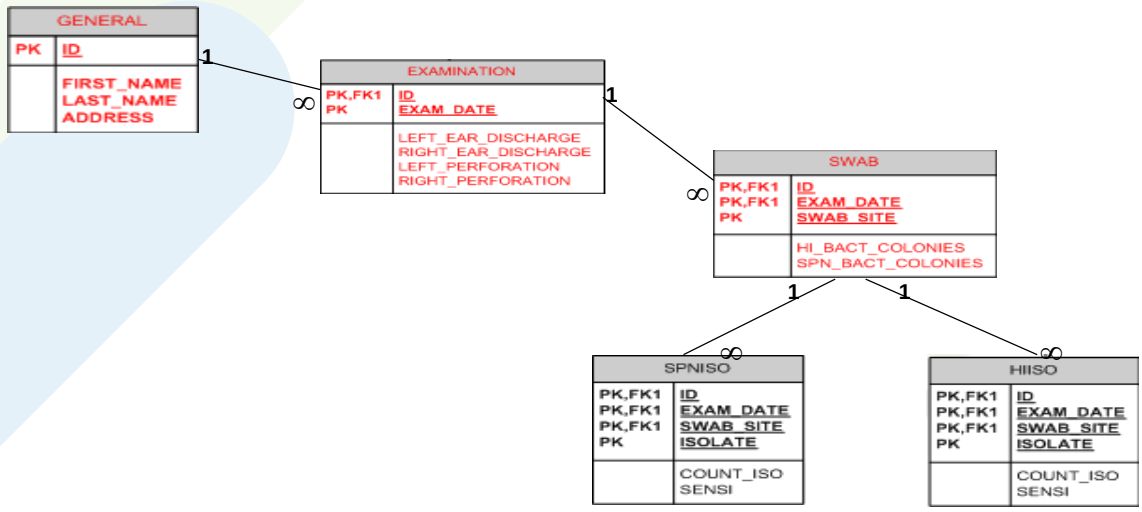
ID	NAME	DOB	AGE	EXAM_DATE	BP
1	Joe Bloggs	1/03/1987	20	12/05/2007	120/80
2	Smith, Jane	12/05/1998		12/05/2007	110/70
1	Joe Blogs	1/03/1978	21	1/06/2008	130/90

ID	EXAM_DATE	MEDICATION	DOSE
1	12/05/2007	Penecillin	400 mg daily for 7 days
2	12/05/2007	Amoxycillin	350
1	1/06/2008	Penicilin	200 mg twice a day for 7 days

46

46

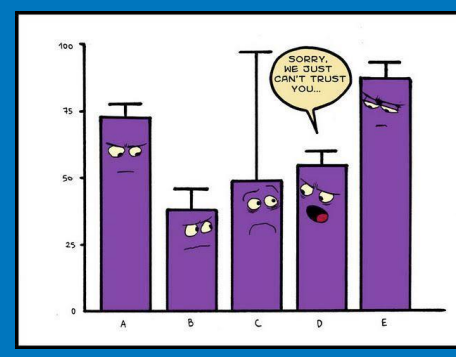
Relational Database



47

47

Minimising Data Errors



48

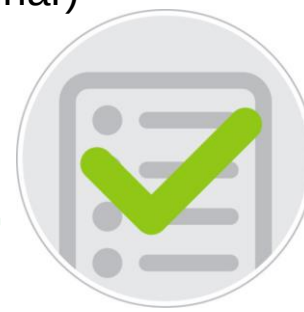
48

Minimising Data Errors

You now have a suitable database structure

Think further about maximising data integrity ***before analysis***

- Instrument design (see Survey Design seminar)
- Database design
- Testing the Database
- Data entry
- Validation
- Cleaning

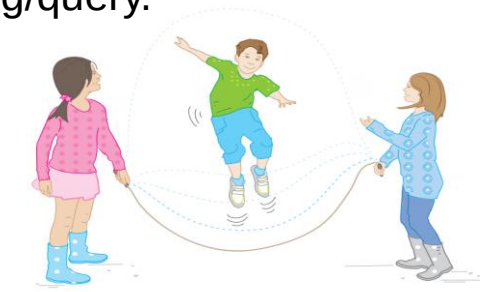


49

49

Database Design

- Minimise data errors
- Specify only certain values to be entered into a field
  - legal values for categorical data, e.g. 0, 1, 7 or 9
  - range for continuous data and dates, e.g. 75-100.
  - Always include codes for missing/query.
- Compulsory fields
- Unique fields
- Skips (branching logic)



50

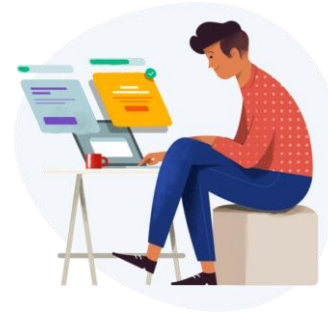
Test your Database

1. Ensure structure and integrity checks work as planned
2. Try to “crash” the database
3. Verify nothing unexpected
4. *ensure error messages/ prompts if incorrect data entered

51

51

How to Test a Database



- Try entering a duplicate record
- Test skips work
- Check required fields cannot be left blank
- Test warnings for attempts at incorrect values
- Categorical variables: test only valid responses work
- Continuous variables: test max, min and beyond
- Logic checks *mouse versus enter/tab/arrow etc.

52

52

Data Entry Choices

- Manual or scanned
- Consider cost, accuracy
- How to make it better?
  - Standard Operating Procedures
  - Training
  - Regular review in real-time

53

53

Data Entry Validation

- Double entry e.g. EpiData
- Hard copy vs electronic copy checks – 10%

54

54

Concurrent Data Entry

- Some packages allow multiple users: eg REDCap
- EpiData is single user but can append multiple files
- May be best managed by always entering into a blank dataset → one person responsible for merging

55

55

Data cleaning

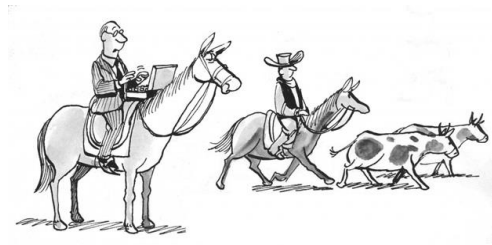
- Check no query codes are left and…..
- CLOSE THE DATABASE → cleaning
- Allow time, <u>lots of time</u>
- **NEVER** change the original dataset
  – use an analysis/command file – 1st step "save as"
- Recode/modify variables only in command files
- Exception: recoding to missing values
- Must be documented and reproducible

56

56

Missing Data

- Check for blanks – should only be for skips
- Errors should be recoded to the missing code
- When all missing data are best accounted for: recode to the program value for missing data



"Ooo, two strays to add to the database."

57

57

Logical Checks prior to Analysis

- Plausible values
- Inconsistency checks
- Find duplicate records
- Check eligibility criteria

58

58

Cleaning Categorical Variables

- Produce frequency tables for each variable, OR
- Explicit checks on values
  - Assert alive = 0, 1, 9 (if fails then errors exist)
  - Checking there are no query codes remaining

"This is not what I meant when I said 'we need better data cleansing!'"

59

59

Cleaning Continuous Variables

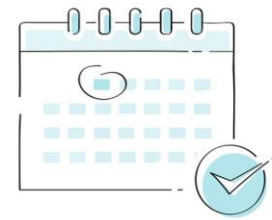
- Specify reasonable upper and lower limits
  - Check outliers and correct
  - Replace as “missing” if impossible value
- For each variable produce summaries of
  - mean, median, variance, max and minimum
- Or use a dot plot to easily spot possible errors

60

60

Cleaning dates

- Check dates in correct order
  - Dob< date of 1st visit < date of 2nd visit
- Check reasonable time spans
- Calculate ages, time intervals and look for implausibles
  - E.g. negative age



61

Example 1

alive	Freq.	Percent	Cum.
1	66	54.55	54.55
2	54	44.63	99.17
3	1	0.83	100.00
Total	121	100.00	

62

62

Example 2

```
     Variable |        Obs        Mean    Std. Dev.
-------------+--------------------------------------
      weight2 |        121    26.00826    32.25148
```

63

Example 3

Once the missing codes are converted to missing values, the mean and standard deviation are much smaller. The number of observations is also reduced.

Check the min/max values are within the plausible range.

```
  Variable |    Obs        Mean    Std. Dev.       Min        Max
-----------+---------------------------------------------------------
   weight2 |    101    11.73267    1.702299         9         19
```

64

Example 4

age2 = age at 2 year visit (yrs)

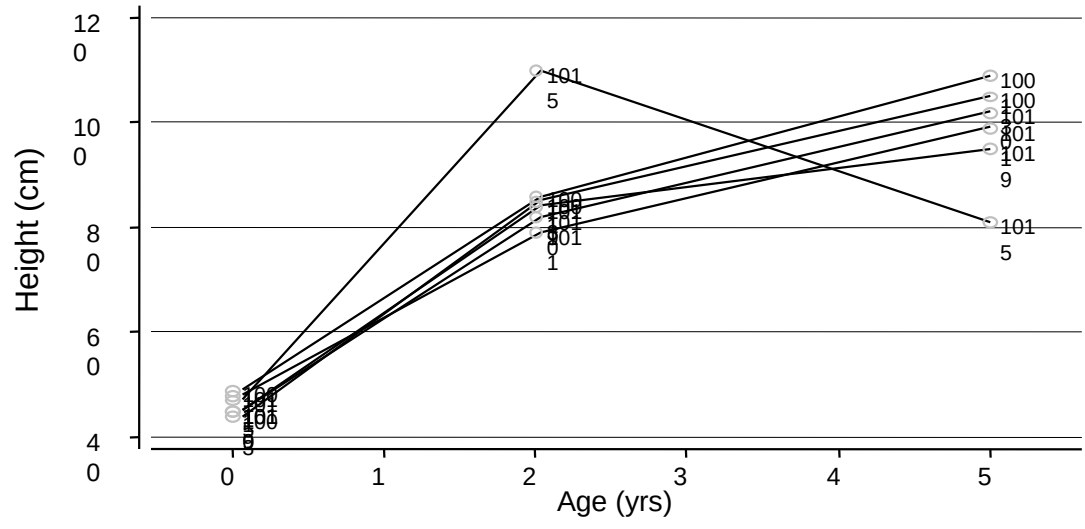
Calculated from datebir and date2yr

Variable	Obs	Mean	Std. Dev.	Min	Max
age2	103	2.2297	5.193444	-24.9911	47.02806

65

65

Example 5



66

66

Resources

67

Data Managers

- Develop/fine tune data collection form
- Database design
- Data dictionary creation
- Database creation
- Database queries and reports
- Security/access to the database
- Advice on data cleaning

"Guide us, Oh Database Manager!"

Collaborate!

Get advice early and often

68

REDCap Access at CAHS

- Licensing conditions mean projects must have a TKI employee to reside on the Telethon Kids instance of REDCap
- All existing projects set up on TKI REDCap can remain short-mid term
- Dept of Health REDCap now available for all WA Health employees with an HE number and health email address
- See handout for more details

69

69

REDCap resources

REDCap
Research Electronic Data Capture

CAHS Research Education Program

- Workshops + handout resources
- Overview seminar
- Training videos
- Access and supports
- https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Past-seminars
- https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Workshops

70

70

Other Resources – see Handout

e.g:

- CAHS.ResearchSupport@health.wa.gov.au - ✉
- Telethon Kids biostatistics@telethonkids.org.au - ✉
- Biostatisticians and data managers everywhere
- REDCap https://www.project-redcap.org/ - 🌏
- Good Clinical Practice training - 🌏
  - GCP Mutual Recognition (transcelerate-gcp-mutual-recognition.com)
  - ICH Good Clinical Practice E6 (R2) • Global Health Training Centre (tghn.org)
  - Online Training - RETProgram

71

71

Summary: Steps to Good Data Management

- Use well-designed data collection forms
- Use an appropriate data entry package
- Have a data management plan
- Create a data dictionary
- Use a relational database
- Test your database
- Carry out data cleaning before analysis
- Keep a record of all analyses
- Archive all data at the completion of the project

And don't forget to budget for these activities!

72

72

Take Home Messages

- It's all about planning
- Get advice early and often
- Data management requires you to be
  - Meticulous
  - Obsessive
  - Systematic
  - Logical
- *Garbage in = Garbage out

"The difference between something good and something great is attention to detail." *Charles R Swindoll*

73

73

Upcoming Research Skills Seminars

30 Jul **Consumer and Community** with **Anne McKenzie AM**

6 Aug **Knowledge Translation** with **Dr Fenella Gill**

13 Aug **Media and Communications** with **Elizabeth Chester**

Register → **https://researcheducationprogram.eventbrite.com.au**

We love feedback

A survey is included in the back of your handout or complete online via: https://tinyurl.com/datacollandmgmnt

☎ (08) 64565014 ✉ ResearchEducationProgram@health.wa.gov.au cahs.health.wa.gov.au/ResearchEducationProgram

74

74

Government of Western Australia
Child and Adolescent Health Service

© 2021 CAHS Research Education Program

Child and Adolescent Health Service Department of Research
Department of Health, Government of Western Australia

☎ (08) 6456 5014 ✉ ResearchEducationProgram@health.wa.gov.au
cahs.health.wa.gov.au/ResearchEducationProgram

75

Government of **Western Australia**
Child and Adolescent Health Service

Data Collection and Management

RESOURCE NOTES

CAHS Research Education Program
Research Skills Seminar Series
☎ (08) 6456 4585 ✉ ResearchEducationProgram@health.wa.gov.au
🌏 https://cahs.health.wa.gov.au/ResearchEducationProgram

Table of Contents

1. Why we need good data management

The whole point of conducting research is to obtain high quality data that can have impact. This is an ethical issue, as poor quality, corrupted or lost data can mean not answering the research question, or falsely influencing policy and practice – in other words, potentially wasting participants and your time and resources, and affecting your reputation as well as that of your institution.

Good data management is a foundation of good research. It should be planned from the beginning of a research project and become part of standard research practice. Unfortunately, data management is often done at the last minute, using the first method that comes to mind. This approach is generally more time-consuming and error-prone. Taking time at the start of a research project to put in place robust, easy-to-use data management procedures will usually pay off several times over in the later stages of the project. If data are properly organised, preserved and well documented, and their accuracy, validity and integrity is controlled at all times, the result is high quality data, efficient research, outputs based on solid evidence and the saving of time and resources. By contrast, inadequate data management can lead to catastrophes like the loss of data or the violation of people's privacy. Some journals also require data to be entered to a data repository for open access, or availability of data on request.

In short, research data should be accurate, complete, identifiable, securely stored, retrievable and able to be made available to others. Good management of data results in high quality data able to provide a meaningful and interpretable result for a research question.

2. What are the steps to achieving good data management?

Data management includes all activities associated with data other than the direct use of the data. Steps to be considered are:

- Analysis planning – based on a clear, answerable question and objectives
- Data base design and development
- Database testing
- Developing a data management plan
- Data file management
- Coding sheets and Data dictionaries
- Data collection
- Data entry
- Data validation
- Data cleaning
- Security
- Data sharing and collaboration
- Archiving

2.1. Research Data Management - Good Sources of Open Access Learning Materials

- **Australian Research Data Commons (ARDC)**
  Webinar – Data Management in the revised National Statement on Ethical Conduct in Human Research - 5 September 2018
  https://register.gotowebinar.com/register/6111782010487118338

- **Curtin University Library - Research Data Management**
  http://libguides.library.curtin.edu.au/research-data-management

- **Research Data Services**
  https://www.rds.edu.au

- **Australian National Data Service (ANDS) Data Management**
  http://www.ands.org.au/working-with-data/data-management

- **Edith Cowan University**
  - Research Journey for Research Students – Data Management
    http://intranet.ecu.edu.au/research/for-research-students/research-journey/designing-and-undertaking-your-research/data-management
  - Library Guides – Manage Research Data
    http://ecu.au.libguides.com/research-data-management

- **Coursera online courses**
  https://www.coursera.org/courses?query=data%20management
  - Data Management for Clinical Research
    https://www.coursera.org/learn/clinical-data-management

- **Zenodo Research data management (RDM) open training materials**
  https://zenodo.org/communities/dcc-rdm-training-materials/?page=1&size=20

- **University of Oxford - IT Services Research Support Team**
  http://blogs.it.ox.ac.uk/acit-rs-team/advice/research-data-management/rdm-training-resources/

- **Digital Curation Centre**
  http://www.dcc.ac.uk/

- **UK Data Archive**
  http://www.data-archive.ac.uk

3. Data management responsibilities

Funding bodies and governments increasingly require sound data management. Researchers have a responsibility to make themselves aware of any relevant codes and to comply with them. Failure to comply with funding body requirements (eg In Australia: ARC or NHMRC – increasingly required in grant applications – so PLAN, get input from IT, data services etc) may jeopardise future research funding. Failure to comply with legal requirements, such as those that safeguard the privacy of participants in medical research, may lead to prosecution.

Good Clinical Practice training provides information on many aspects of standards required for data collection and management in clinical research.

“An introduction to GCP” one hour overview is available via the Research Education Program at:

https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Past-seminars

Training options include:

- **Global Health Trials**
  https://globalhealthtrials.tghn.org/elearning/

- **ARCS Australia**
  https://www.arcs.com.au/events/category/online-learning

- **Research Education & Training Program (RETProgram) WAHTN**
  https://retprogram.org/portfolio-item/ich-good-clinical-practice-gcp-e6-r2/

The state public sector in Western Australia does not currently have a legislative privacy regime. Various confidentiality provisions cover government agencies and some of the privacy principles are provided for in the Freedom of Information Act 1992 (WA).

http://www.austlii.edu.au/au/legis/wa/consol_act/foia1992222/

3.1. Important documents/sites

- WA Health Research. Governance Policy and Procedures Handbook.
  http://www.health.wa.gov.au/CircularsNew/attachments/724.pdf

- WA Health Patient Information Retention and Disposal Schedule (PIRDS) 2016
  http://www.health.wa.gov.au/circularsnew/circular.cfm?Circ_ID=13308

- Research Data Management Toolkit. University of Western Australia.
  Includes Planning, Intellectual Property, Documentation, Storage/Backup, Sharing/Reuse, Retention/Disposal, Support/Contacts/Useful Resources.
  https://guides.library.uwa.edu.au/RDMtoolkit

- The Telethon Kids Institute has policies for Confidentiality of Research Data, Information Retention and Disposal, Archiving, and Information Security and Handling Procedures.

- Souhami R. Governance of research that uses identifiable personal data. BMJ. http://www.bmj.com/content/333/7563/315

- Australian Code for the Responsible Conduct of Research (2018) https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018

- WA Health and Institutional policies – Research Governance Framework https://rgs.health.wa.gov.au/Pages/Research-Governance-Framework.aspx

- UWA Code of Conduct for the Responsible Practice of Research. Section 2. University of Western Australia http://www.governance.uwa.edu.au/procedures/policies/policies-and-procedures?method=document&id=UP12/25

- Australian Clinical Trials Handbook https://www.tga.gov.au/publication/australian-clinical-trial-handbook

- Information Lifecycle Management System 0557/14 http://www.health.wa.gov.au/circularsnew/circular.cfm?Circ_ID=13145

- WA Health Portable Computer and Storage Devices Policy 2009 http://www.health.wa.gov.au/CircularsNew/attachments/397.pdf

- WA Health Information Storage and Disposal Policy Sep 14 http://www.health.wa.gov.au/circularsnew/attachments/946.pdf

- WA Health Hospital Morbidity Data System. HMDS Reference Manual July 2014. https://ww2.health.wa.gov.au/-/media/Files/Corporate/general-documents/Clinical-Information-Assurance/Part-A-HMDS-Ref-Manual-2018-19.pdf

4. Data management plans

Documentation of project information in a Data Management Plan is an invaluable resource to later members of the project team as well as other researchers that come to investigate your project and its data after project completion, or audit personnel. Increasingly these are required for grant applications. Get input from IT, data services etc.

A data management plan needs to cover:

1. Survey of existing data: What existing data will need to be managed?
2. Data to be created: What new data will your project create?
3. Data owners & stakeholders: Who will own the data created, and who would be interested in it?
4. File formats: What file formats will you use for your data?
5. Metadata: What metadata will you keep? What format or standard will you follow?
6. Access & security: Who will have access to your data? If the data is sensitive, how will you protect it from unauthorised access?
7. Data organisation: How will you name your data files? How will you organise your data into folders? How will you manage transfers and synchronisation of data between different machines? How will you manage collaborative writing with your colleagues? How will you keep track of the different versions of your data files and documents?
8. Storage: Where will your data be stored? Who will pay? Who will manage it?
9. Backups: Hard drives on desktop and laptop computers fail regularly. You must follow a credible backup strategy of regular backups. Consider including an off-site backup so that your data will not be lost in a “worst case scenario” eg your building burns down. Rather than relying on memory, consider automated backup.
10. Bibliography management: What bibliography management tools will you use? How will you share references with the other members of your group?
11. Data sharing, publishing and archiving: What data will you share with others? Who will be allowed to have future access or re-use data? How will you organise this?
12. Retention and disposal procedures and provisions: What data will you destroy? When? How?
13. Responsibilities: Who will be responsible for each of the items in this plan?
14. Budget: What will this plan cost? Possible costs include those for backups, research assistant time for data curation, metadata creation, archiving, data manager time etc.
15. Ownership and protection of intellectual property
16. Anything else: Don't restrict yourself to the items above. Stop and think. Is there anything missing?

5. Good file management practices

Standardisation creates uniformity within/across projects and allows for easy retrieval of files because everyone knows where to find them, as well as version control. Covered in more detail by Good Clinical Practice (GCP) training.

A few basics:

- Maintain a single master copy of your data file.
- Always make a copy of the data file with a new name before any data entry or edits are made.
- Name your copies using the current date e.g. MyDataFile_2018-09-28.rec. This y-m-d format helps you sort files in date order.

- Define an appropriate directory structure that makes clear what the purpose is of any files therein: e.g.
  /myproject
  /myproject/data
  /myproject/data/backups

6. Confidentiality

- Keep the database and/or your computer password protected.
- **Use codes to de-identify the data**, i.e. id numbers (never an "identifiable" number such as a hospital record number)
- Keep the participant's name and contact details in a separate file to their survey results. For longitudinal studies, details such as contact details of family/friends, withdrawals should also be "stripped" from the main database and kept separately
- Ideally use a data capture program like REDCap which enables
  - de-identification of identifying variables (e.g. Name, date of birth, hospital record number, contact details), and enforces allocation of a de-identied "unique identifier – that can be used to link back to identifying details if required
  - database users to be ascribed user rights at different levels – i.e. only certain people will have access to data, and the level of access can be controlled
- Ensure hard copy data is kept in locked filing cabinet (or equivalent).
- In WA Health, de-identified data may be moved on a Dept of Health encrypted USB but NOT stored.
- Have a standard operating procedure to ensure all project staff understand their responsibilities around keeping data confidentially.
- Ensure you are familiar with your institution/state/national policies on data confidentiality – this includes movement of and access to data
- See websites earlier in this handout for further information and policy.

7. Data sharing / collaborative data entry work

- REDCap is a good example of data capture software enabling multiple users, with the ability to keep certain areas accessible only to certain people, and multiple data entry persons.
- If using a server-based database for multiple users, ensure only the latest version can be accessed
- If you don't have access to a joint access database, concurrent data entry may be best managed by always entering data into a blank dataset, then having one person responsible for merging the results
- If a single-user database, each user can be given a separate copy of the database with its own name into which they enter all their data. The data in these separate copies can later be appended. Note that this remains less optimal to a multi-user data entry system like REDCAP as appending will not prevent a duplicate record from being entered – e.g. one by a user into their copy of the database and another by a second user into their copy.

- UK Data Archives - excellent Guide To Managing and Sharing Data May 2011 http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
- UWA Research Data Sharing/Re-use (Part of the Research Data Management Tool Kit) http://guides.is.uwa.edu.au/content.php?pid=319161&sid=2612329

8. Data archiving / storage and data destruction

At the end of the project create an archive of:

- All data, electronic and paper
- All cleaning and analysis command files
- All related documentation
- Date when the archive was created and know when or if it can be destroyed
- In WA, minimum storage times are 5 years after all reference to the documents has ceased (data or related research documents) AND for interventional studies involving children, until the child reaches age 25 years. Clinical trials data are generally kept for 15 years.
- Hard copies can be archived off site as long as they are retrievable for the required period and are not required frequently (e.g. not more than once per year)

No destruction of records, information or data can be conducted unless it is in accordance with an approved disposal authority;

The approved disposal authorities which should be used for WA Health staff are contained within

- The Patient Information Schedule RD 2007079;
- Administrative and Functional Records RD 2005001 (available from the Department of Health) and
- The General Disposal Authority for State Information (available from our website – www.sro.wa.gov.au);

All information and data able to be legally destroyed now, must be documented – that is all records, information and data which can legally be destroyed must be listed with proof of the authorised disposal authority under which it is being destroyed.

At a minimum, this information must include:

- file or document group or subject;
- date range of same;
- identification of the disposal authority – name and number;
- evidence of sign-off by CEO;
- date of destruction; and
- method of destruction.

For the destruction of data and electronic records, the metadata must be kept which identifies the items listed above.

As the State Archive is full, the Department of Health advises suitable alternative storage via :
WA Dept of Health 2014: Data Storage and Disposal Policy Aug 2014
http://www.health.wa.gov.au/circularsnew/attachments/946.pdf

Note there are two main types of data to be considered: type 1 below should be retained, type 2 can eventually be destroyed after following the appropriate timing recommendations. If in doubt, keep it.

8.1. Research administrative and functional records (approval, monitoring, publications etc)

Retain

- Records associated with the activities involved in investigating or enquiring into a subject or area of interest in order to discover facts, principles etc. e.g. Ethics and governance documentation

8.2. Patient information (data, consent etc)

Eventual Destruction

- Patient or Subject Research Records - Refers to records (including laboratory results, reports, questionnaires and surveys) obtained from consenting patients or subjects for the specific purpose of researching a project, theory or trial.
- Records of Consent or Authorisation - Refers to records of consent or authorisation for the use of patient's or subject's results in research.
- Research Requests - Refers to documented requests to perform research.

8.3. Preparation for storage - courtesy of State Records Office

- Group the records into their proper series (SRO can assist here if you're unsure).
- Box the records sequentially by their numeric, alphabetic or alpha-numeric numbering system.
- List the records in full on consignment/manifest lists in the same order in which they are boxed (i.e. sequentially).
- Include a copy of the consignment list in the box and record details in your organization's Recordkeeping system.
- Ensure boxes are numbered sequentially and the range of items for each box is noted on the end of the box (i.e. the first and last record number for each box).
- De-metalling (removal of staples, paper clips, metal fasteners) is not recommended when preparing archives for transfer

Dept of Health WA recommends archival records awaiting transfer to the SRO/recommended facility be boxed by the custodial organisation in approved archive boxes. Standard archive boxes are made from acid free board (neutral pH and buffered).

The standard box, suitable for most files, measures 385mm x 250mm x 168mm (internal measurements when box is made up). These boxes are commonly referred to as Type 1 archive boxes.

To determine how many boxes will be needed, measure the length of the files as they would stand on a shelf and divide by 0.16m. Other size boxes are also available to accommodate smaller or larger items. If not available at the Health Dept., archival quality

boxes can be ordered from suppliers.

Some agencies have needed to utilise commercial storage providers on the current Common Use Arrangement:
https://www.contractswa.finance.wa.gov.au/group.jsp?groupID=OF&STMP=14052015024 8112#17

9. Data collection and analysis planning

9.1. Basics of setting up databases

9.1.1. What is a database?

- A computer software program that facilitates the entry, storage and manipulation of data
- The data is stored in a table and the file is called a data file
- Each row represents a case/record
- Each column represents a field/variable
- One or more of these tables stored together create a database.
- A relational database has multiple tables that are related to each other. Usually they are linked together by unique identifiers.

9.2. Database software

9.2.1. REDCap – the preferred data capture program for WA Health and many leading institutions

- Data stored locally, accessed via web
- FREE – acknowledge only
- Online tutorials available
- Standard forms provide assistance with set up
- Front end checks and balances available to improve accuracy of data entry
- See extensive information on the Research Education Program open access website
- Seminar 1h overview (section 6): https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Past-seminars
- REDCap Workshops and associated handouts https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Workshops
- General access and help information including other instructional videos: https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Workshops

9.2.2. EpiData

- Small free portable program (www.epidata.dk)
- Data entry forms can be set up to resemble paper questionnaire

9.2.3. SPSS Data Entry

- Stand-alone product that will allow validation checks

9.2.4. Medrio

- for clinical trial data entry – cloud-based, secure

9.2.5. Webspirit

- For clinical trial data – available through Paediatric Trial Network Australia

9.2.6. Qualtrics

- Paid online tool for creating surveys (www.qualtrics.com)
- used by a number of institutions in WA eg Curtin, UWA, the Raine Study

9.2.7. Microsoft Access

- Has limitations – security issues, potential to overwrite data etc
- Can deal with complex relational databases
- Data entry forms can be set up to look like your paper questionnaire
- Able to support multi user access to database
- Customised reporting available

9.2.8. Microsoft Excel - users BEWARE

- NOT recommended. VERY EASY TO DESTROY YOUR DATA IRREVOCABLY
- Comprehensive data checking will be required once all the data have been entered
- Unable to enforce uniqueness for an identifier
- Not a relational database
- Only one person can access a file at any one time.
- Need to be careful with dates

9.2.9. Survey Monkey - users BEWARE

- NOT recommended. Preferably don't use, or use for non-sensitive data with caution
- You don't own the data
- Unclear where data are
- You may be breaching Dept of Health policy
- Never use for sensitive, identifiable or re-identifiable health data
- **Not recommended by many sites** including Telethon Kids Institute and Perth Children's Hospital for research data.
- **Evolving policy to watch for at WA Health** – not recommended for patient or staff data, particularly if potentially sensitive

9.3. Variables, coding sheets and data dictionaries

Before creating a database, you need to design its structure. This is done by creating a data dictionary or coding manual in conjunction with the data collection forms.

A data dictionary includes:

- Table names
- Variable names
- Variable descriptions
  - meaning, data type, units of measure

- Validation/coding rules
  - code for categorical variables
  - ranges for continuous data and dates
  - codes for missing data
- Relationships between tables

9.3.1. Naming Variables

- Each variable must have a unique name
- Choose an informative name – something anyone could understand
- Do not use spaces, special characters or punctuation marks
- Begin variable names with a letter
  Select variable names that are compatible with both your data entry and statistical packages.

  Check:
  - The maximum number of characters allowed for a variable name
  - The characters that can be used to make up a variable name
  - Is the package case sensitive?
  - Does the package have any special names that can't be used for variable names?

9.3.2. Examples of statistical package variable naming issues

- Variable names that start with a number will be prefixed by an "n" in EpiData.
- The underscore character "_" is OK in Stata but not in EpiData.
- Lowercase is recommended if you will be analysing your data in Stata.
- In EpiData a variable name can be up to 10 characters long, contain letters or numbers, and must begin with a letter.
- REDCap variables may not start with a number, use underscores not spaces, and can contain lowercase letters or numbers
- If your variable names contain numbers only, Stata will name the variables v#.
- If your variable names start with a number, the number will be stripped in Stata.
- MySQL and Oracle databases will not accept variable names starting with a number.

9.3.3. Selecting data types

- Data are generally numeric, text or date
- Use numeric variables for continuous data such as height and weight or for variables on which you wish to perform mathematical operations.
- Use numeric codes to hold categorical data wherever possible – easier for data entry plus many statistical procedures will only work with data stored in numeric format.
- If using string codes (free text), be consistent with spelling and upper vs. lower case text.
- For date variables select a date format rather than entering dates as strings or entering day, month and year as separate variables (otherwise very hard to work with).

9.3.4. Numeric Data

Categorical

- Nominal
  - no ordering implied
  - binary if only 2 categories
  - no inherent meaning e.g.: 1=single; 2=married; 3=defacto; 4=divorced/separated
- Ordinal
  - categories assume a natural ordering
  - codes convey the order e.g: 1=poor; 2=fair; 3=good; 4=very good; 5=excellent
- Continuous
  - e.g. height, weight, waist
  - – can only take specific values (visit #: 1 2 3 4…)

String/text

- can be categorical, e.g. never; sometimes; always
- you will need to convert string variables to numeric formats in your statistical package prior to doing most statistical procedures, e.g. never=1, sometimes = 2, always = 3
- generally avoid wherever possible

Date

- Use of a date format in database packages such as REDCap, EpiData and Access will ensure that only valid dates will be accepted.

9.3.5. Continuous and discrete variables

Continuous and discrete variables contain numbers only, for example height or weight (continuous) or number of people in a household (discrete). You will often have limits or ranges for continuous variable. For example, an Apgar score must be an integer between 0 and 10. When setting up a database you can specify these limits to reduce the chance of data entry error. Many measurements come in units such as days, grams or millilitres. Decide on the most appropriate unit (grams or kilograms? minutes or hours?) and specify this unit on the coding sheet, questionnaire and data entry form. All entries for one variable must use the same unit. Don't mix up (for example) grams and kilograms or months and years. Setting ranges for variables helps prevent this.

9.3.6. String variables

String variables can contain text and numbers. As it is difficult to perform mathematical or statistical operations on strings, they should not be used if a numeric variable can be used instead. You might, for instance, have a question that asks for community. The paper questionnaire can have a space for the community to be filled in, but before the questionnaires are entered in the database, you can give each community a numeric code which is entered in the database. String variables usually contain data that is too complicated to categorise, such as long comments. Many programs have limitations on the length of string variables.

9.3.7. Date variables

Date formats vary around the world. Both day-month-year and year-month-day are logical date formats that are easily understood. Use one date format that will be understood by everybody working with your data, and use it consistently throughout the database.

9.3.8. Unique identifiers

- Each record or case must have a unique identifier
- Unique identifiers are used to refer to specific records in the database *without using identifying information* such as names. They are also used to link data between tables in the database. The unique identifier for a participant is usually a number assigned by the researcher.
- The unique identifier can be a single field, such as studyid, or a combination of two or more fields, e.g. studyid and visitdate.
- Hospital record number should not be used as the unique identifier.
- If using REDCap, the first variable will automatically be set up as the unique identifier, and will generally be consecutive numbers.

9.3.9. Missing data

- Include a code for missing data
- For numeric variables, missing data is conventionally represented by “9” for one-digit variables, “99” for two-digit fields, and so on. The missing code must never be a valid response for that variable. For dates, 9/9/9999 can often be used as the missing code.
- Some fields, such as id numbers and eligibility criteria, should never be missing.
- Change missing codes to missing values in your statistical package before you begin your analysis.

Using special codes for missing data allows data entry personnel to indicate a blank on the data collection form as opposed to data that has been accidentally skipped during data entry. Fields with missing data can be collated later on and efforts made to retrieve missing data where possible. All fields should have data in them unless they are skipped because of a conditional jump. For example, if a participant says no to “Have you been diagnosed as diabetic?”, then they shouldn’t have data for the question “Date of diabetes diagnosis”.

*The missing code must never be a valid response for that variable, so if someone’s weight could be 99, then “999”, and not “99”, should be used as a missing code. Alternatively to maintain consistency you may prefer to nominate a large number, say 9999, as the missing code for all your numeric variables.

9.3.10. Data Entry Queries

- Include a code for data queries encountered during data entry
  For example, if a response written on a data collection form is illegible, unlikely, unclear or inconsistent with other responses then the data entry operator can enter a query code. This indicates something that needs to be followed up.
- Use a value that is not a valid response for the field to indicate a query (e.g. 77 or 7/7/7777 for dates).
- Unique identifier fields must have a valid value and therefore should not have a query code.
- All queries should be dealt with before beginning analysis.

Examples:
- Date of birth is not feasible
- Date field is outside the date range for the project.
- Male respondent answers yes to "Have you been pregnant?"

Consistent dates and times

Example of data downloaded from hospital patient admission system. The date of admission and date of discharge were exported from the hospital system formatted as mm/dd/yyyy and emailed to the project coordinator. Not realising the formatting issue, the data was imported in to the project database which used dates formatted as dd/mm/yyyy. The problem arose when only the dates with days greater than 12 were transposed to dd/mm/yyyy format and the rest were unchanged. This of course meant that many of the dates were incorrect. In this case the data had to be requested again specifying the dates to be formatted as dd/mm/yyyy.

Enter raw data rather than summary data

e.g. when measuring heights of 2-year olds you may take three measurements and use the average of the three measurements in analysis. Still enter all three measurements and allow your stats package to calculate the average. Doing this also allows you to check how discrepant the three measurements are. If one is very different from the other two you may want to calculate the average using the other two measurements only. Entering average height: you're more likely to make an error calculating the average than your stats package will, and it's more difficult to check the computerized records against the paper records.

Do not confuse data entry and calculations.

Calculations can be done later in your statistical package. The exception would be where the results of a calculation would inform your next steps, e.g. how to treat a patient, determining what other data needs to be collected, etc. Sometimes calculated fields are included on the data entry form for data validation but these fields are not saved to the database.

Identify sections within a data collection form with repeating information collected on different subjects or at different time points.

For instance, assume we're collecting data on pregnant women and the babies they deliver. A pregnancy can result in the birth of more than one baby. In the database how many babies should we allow for? Singletons, twins, triplets, quads? We could create one file with enough space to allow for quads but this would result in a very

wide file with many empty variables. An alternative is to enter the data in two separate data files, one for pregnancy and another for babies. The baby file should include the pregnancy id number so that each baby can be linked to the pregnancy. There would be a separate table for each mother and a mother may have several pregnancies over the course of the data collection phase.

9.3.11. Badly Designed Table – Example 1

ID	NAME	DOB	AGE	EXAM_DATE	BP
1	Joe Bloggs	1/03/1987	20	12/05/2007	120/80
2	Smith, Jane	12/05/1998		12/05/2007	110/70
1	Joe Blogs	1/03/1978	21	1/06/2008	130/90

- Table holds data on more than one subject - demographics and examinations
- NAME and BP fields are not atomic – break into FIRST_NAME and SURNAME fields, SYS_BP and DIAS_BP
- AGE field has not been calculated by the computer program and is more likely to result in errors.

Calculated fields are not stored in the database because they can lead to inconsistencies, for example if the DOB field was updated then the researcher must remember to also update the AGE field. The AGE field could be added on the data entry form and calculated from the DOB and EXAM_DATE fields. This would allow the data entry person to check that the participant is eligible for the study or highlight data entry errors with the date fields.

9.3.12. Badly Designed Table Example 2

ID	EXAM_DATE	MEDICATION	DOSE
1	12/05/2007	Penecillin	400 mg daily for 7 days
2	12/05/2007	Amoxycillin	350
1	1/06/2008	Penicilin	200 mg twice a day for 7 days

- MEDICATION has repetition of text – a drop down list or code set is indicated here
- DOSE field is not atomic – free text field impossible to analyse

An example of how a medication table might be designed is given overleaf.

Variable Name	Description	Data Type	Values	
ID	Unique identification number (Risk Factor Study ID)	number		
DATE_PRES	Date drug prescribed	date		
DRUG_NAME	Antibiotic name	number	1	PROCAINE PENICILLIN
			2	AZITHROMYCIN
			3	AMOXYCILLIN 250MG/5ML
			4	METRONIDAZOLE
			5	BENZATHIINE PENICILIN
			7	LA BICILLIN
			8	SOFRADEX
			9	TRIMETHOPRIM
DOSE	Dose of prescribed antibiotics, amount/dose	number		
DOSE_UNIT	Unit of dose (mls/mgs/units/drops)	number	1	mls
			2	mgs
			3	units
			4	drops
			5	topical
			9	not determined/missing
FREQ	Frequency of dose	number		
FREQ_UNIT	Unit of frequency of dose (daily/weekly/ongoing)	number	1	Daily
			2	Weekly
			3	Ongoing
			9	not determined/missing
TOT_DAYS	Total number of days medication given	number		

9.3.13. Variable Definition Examples

Variable Definition Example 1

Studyid

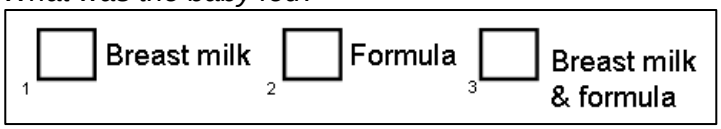
- Essential to have a unique identifier for each record
- Id numbers should be written on every form
- Specify the range of id numbers on the coding sheet
- There should be no missing values

Variable name	Description	Data type	Values/Rules
Studyid	Participant's unique study id number	Number	Must be unique 1001-2000

Variable Definition Example 2

Oralfeed

What was the baby fed?

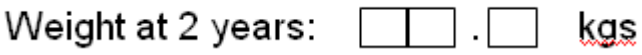


= nominal, categorical variable

Variable name	Description	Data type	Values/Rules
Oralfeed	Oral feeds	Number	1 = breast milk 2 = formula 3 = breast milk & formula 7 = query 9 = missing

Variable Definition Example 3

weight2



- Continuous
- Include unit of measure
- Specify plausible range
- For missing and query codes, use numbers that are outside the possible range of values for this variable

Variable name	Description	Data type	Values/Rules
weight2	Weight at 2 years (kg)	number	< 20 77 = query 99 = missing

9.3.14. Data Dictionary Variable Example - Hospital observation data

ADMISSION OBSERVATIONS

	DATE	Time (hrs)	Temp °C	Heart rate	Resp rate	O_2 Flow (L/min)	SaO_2 on room air
Baseline hospital observations							
Enrolment observations							

WARD OBSERVATIONS UNTIL DISCHARGE

TIME 0 IS THE TIME CLOSEST TO ENROLMENT
PLEASE RECORD OBSERVATIONS EVERY 12 HOURS FROM ENROLMENT ONWARDS UNTIL THE RESPIRATORY EPISODE ENDPOINT (i.e. 16 HOURS OFF OXYGEN, FEEDING ADEQUATELY ETC)

12 HOUR PERIOD	DATE	Time (hrs)	Temp °C	Heart rate	Resp rate	O_2 Flow (L/min)	SaO_2 on room air
1							
2							
3							
4							

This is an example of a data collection form for recording vital observations in hospital. Creating the data dictionary for the form will involve documenting each of the variables including any ranges and code sets.

Creating the data dictionary can be done using Microsoft Word, Excel or some other text file. Each record must relate to a participant who should be allocated a unique identification number. In this case we have created a variable called STUDY_ID for this purpose. The observation time point can be defined as a categorical numeric variable with a code for each observation type (1 = Baseline hospital observation, 2 = Enrolment observation, and 3 = 12 hour period observation).

The rest of the variables have ranges and unit measures as detailed overleaf.

CLIN_OBS Table Clinical observations starting from when the child was admitted to ISOP (7B). Primary key: (study_id, obs_time_pt, obs_date). Relationships: Links to DEMOGRAPHIC table via study_id.			
Variable name	**Description**	**Data Type**	**Values/Rules**
STUDY_ID	Study identification number issued to child from randomisation form. 1000's = Indigenous 26 weeks and less; 2000's = non-Indig 26 wks and less; 3000's = Indig > 26wks; 4000's = non-Indig > 26 weeks	number	1000-4000
OBS_TIME_PT	Clinical observations time point	number	1=Baseline hospital obs
			2=Enrolment obs
			3=12 hourly obs
OBS_DATE	date and time (24hr) of this clinical observation	date	
TEMP	Temperature (deg C)	number	25-45
PULSE	Pulse rate (beats per minute)	number	50-250
RESP	Respiratory rate (breaths per min)	number	20-120
OXY	Supplemental Oxygen (L/min)	number	0-10
RA_SAT	Oxygen saturation on room air (%)	number	60-100

9.3.15. Linking Tables

Tables which need to be linked must contain unique identifiers. It's good practice to give the linking variables the same name in all tables.

9.3.16. Relational Databases

Relational Database Rules (not covered in any detail in the seminar)

Relational database design should adhere to the following rules:

- Each database file should deal with only one project
- Each table should contain data relating to one topic/theme e.g. demographics, lab results, contacts
- Each table must have a primary key/unique identifier, field or fields that are unique for each record
- Each field must relate to the unique topic/theme of the table
- Fields should be atomic, hold just one piece of information
- No derived (calculated) fields in the table
- Repeating groups or fields in a table indicates the need for another table, a one to many relationship (e.g. Med1, med2, med3)

A database design which takes these rules into account allows the data to be easily extracted and manipulated later e.g. REDCap. Excel is NOT a relational database.

Relational Database rules explained

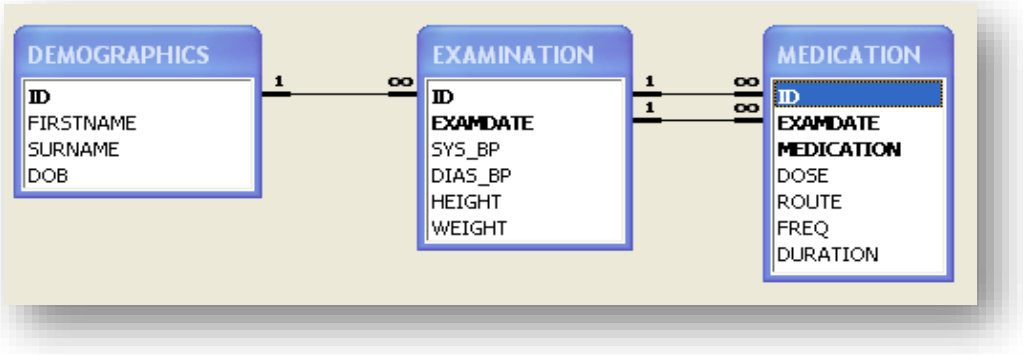
- Each table contains information about one subject. e.g. a "patient" table would contain data at the patient level (dob, gender etc) and a "visit" table would contain data collected at each visit.

- Add a new variable to one table only. Adding "DateOfBirth" to the visit table would result in "DateOfBirth" appearing as many times as there are visits which leads to redundancy and potential data discrepancies.
- Have one piece of information per cell e.g. split patient names into two variables, firstname and lastname.
- Set up all dates and times with the same format, e.g. dd/mm/yyyy, 12 or 24 hour clock.
- Enter raw data rather than summary data
- Repeating information collected on different subjects or at different time points should be stored in a separate table.

Relationship Diagram – Example 1

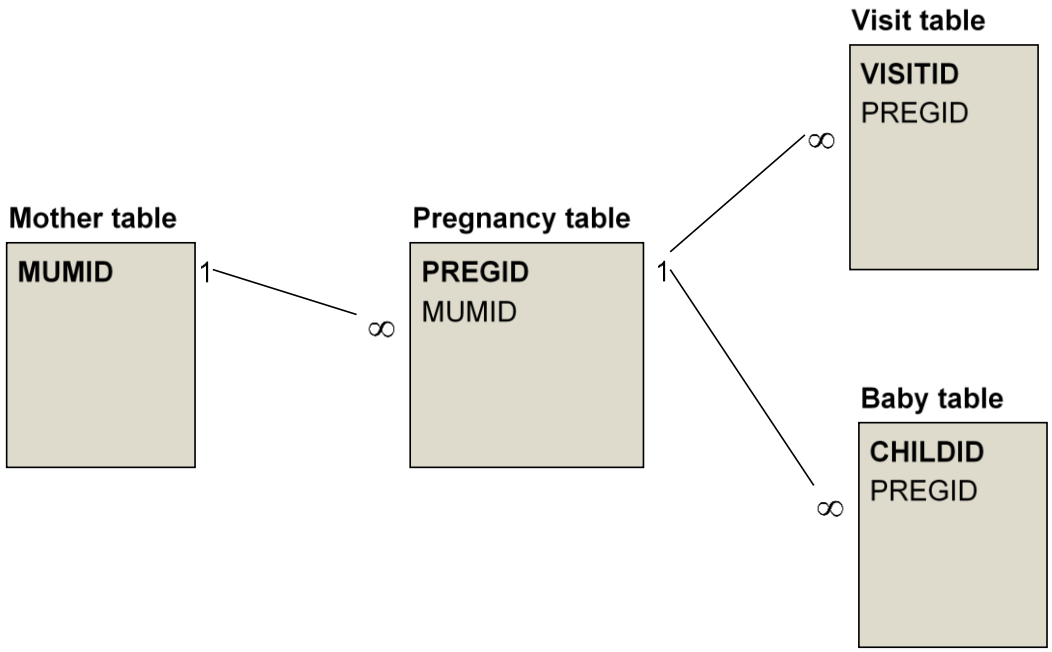
For complex studies, produce a relationship diagram prior to setting up the database and:

- include all the main tables
- indicate the linking fields
- indicate the type of relationship, i.e. one-to-one(1) or one-to-many(∞)



This relationship diagram shows how the previous badly designed table examples (1 and 2 above) could be designed and linked. Both relationships are one-to-many, the person can have more than one examination on different dates and on these dates the person can be given more than one medication.

Relationship Diagram – Example 2



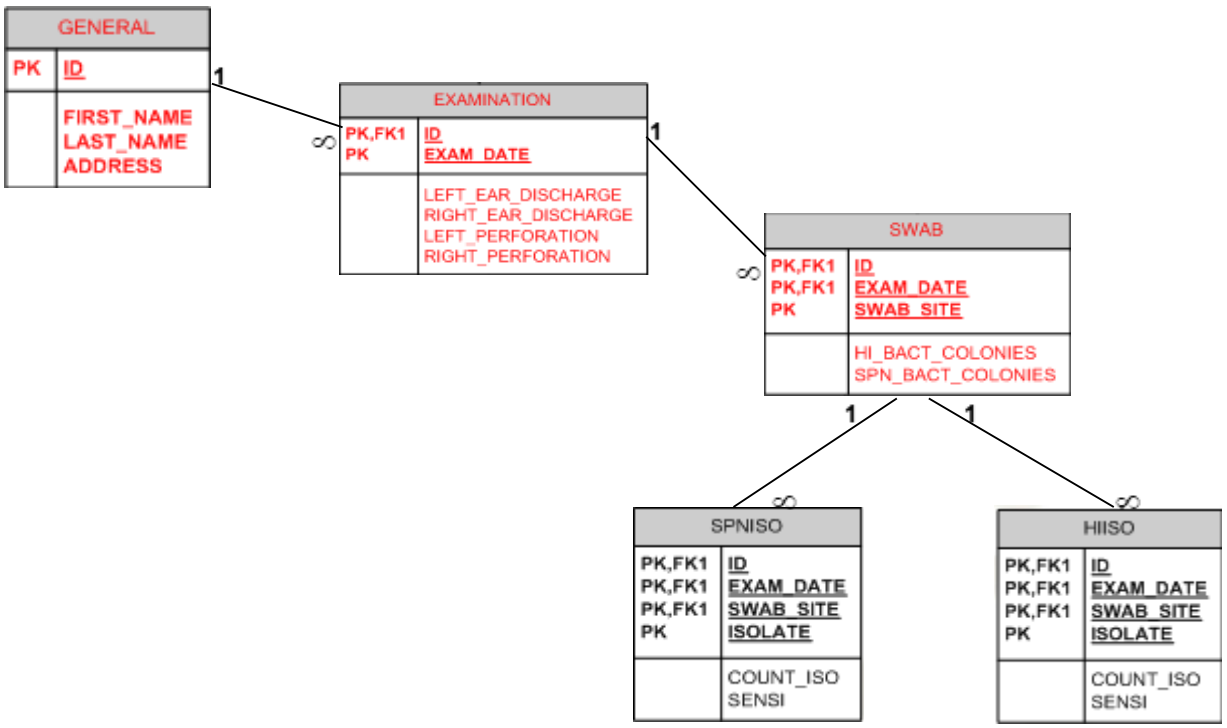
This is an example of how a many-to-many relationship could be represented. If we were collecting data on pregnant women and the babies they deliver, then the mother could have many pregnancies and each pregnancy could result in the birth of more than one baby.

In the database how many babies should we allow for? Singletons, twins, triplets, quads? We could create one table with enough variables to allow for quads but this would result in a very wide file with many empty variables. An alternative is to enter the data in two separate tables, one for pregnancy and another for babies. The baby table should include the pregnancy id number so that each baby can be linked to the pregnancy. There would be a separate table for the mothers and a mother could have several pregnancies over the course of the data collection phase. Therefore we create a junction table with the pregnancy id and the mother id which allows us to link a mother with a pregnancy and a child or children.

This example also shows a visit table where information could be stored on the hospital visits the mother had for each pregnancy.

The unique identifiers for these tables would be MUMID for the mother table, PREG_ID for the pregnancy table, CHILDID for the baby table and VISITID for the visit table. By adding the MUMID into the pregnancy table and PREGID into the visit and baby tables we can link all the information from all four tables.

Relationship Diagram – Example 3



In this example we have a series of one-to-many relationships. The unique identifiers or primary keys (PK) involve more than one field. A participant can have more than one examination, more than one swab can be taken at each examination and these swabs can have more than one isolate extracted from them.

10. Testing a database

The purpose of testing your database is to ensure that it has the structure and integrity checks that you expect. Try to "crash" the database to make sure these checks are working. Verify that the database does everything that you expect it to do, and nothing unexpected.

INSTRUCTIONS

- **Continuous fields**: Test the boundaries of each continuous field by entering minimum and maximum values (should succeed). Then try to enter values just outside the valid range (should fail).
- **Categorical fields**: Check that only valid responses can be entered for categorical fields. Typically a web interface will display only valid values, but EpiData and (sometimes) Access can allow entry of the raw numeric value code. E.g. gender: 1=Male, 2=Female, 8=Query, 9=Missing - ensure these are the only allowed values by testing outside the boundaries: 1, 2, 8, 9 are accepted - 0, 3, 4, 5, 6 and 7 are not.

- **Try entering a duplicate record - and ensure the attempt fails!** Typically this is with a single field, such as a unique study ID number per record in a participant dataset. However in some datasets you may use a combination of fields such as StudyID + Visit Date as a unique "key".
- **Test that skips are working properly**. Consider also what should happen if data is later amended, e.g. what happens to data recorded for pregnancy information when you edit gender from female to male?
- **Check that required fields cannot be left blank**. Note that in EpiData data entry should be done using the Enter, Tab or arrow keys to move from field to field. If using the mouse, jumps and "must enter" rules will be ignored.
- **Test warnings for values that are out of sequence**. e.g. dob should be < visit 1 date, visit 1 date should be < visit 2 date, but can also apply to numeric fields, e.g. systolic bp should be > diastolic bp. For most date fields a date after the current date should not be allowed. Note that "current date" can mean different things: date of questionnaire completion is rarely the same as the date of data entry.
- **Check the time difference between dates is valid**, e.g. if all visits occur between 1 and 2 years of age, check the minimum allowed difference between visit date and dob is 1 year and the maximum allowed difference is 2 years.

11. Data entry

11.1. Strategies for minimising errors

- Use a well designed questionnaire - clear, well presented (see previous Seminar: Survey Design and Techniques)
- Include codes on questionnaire
- Check questionnaires when returned
- Ensure the database fields follow the same sequence as the paper questionnaire
- Set up database to accept only valid responses
- Double enter data
- After data entry clean the entered data (10% check)

11.2. Validation (Database Design)

- Use of validation rules during data entry reduces time spent data cleaning. The following checks are common:
- Allow only certain values to be entered into a field
- Specify legal values for categorical data, e.g. 0, 1, 7 or 9.
- Specify a range for continuous data and dates. Remember the codes for missing/query.
- Program more complex checking procedures, e.g. consistency checks.
- Specify that some fields are compulsory
- Skip fields if particular values are entered
- Specify that id number must be unique

11.3. Double Data Entry

- The idea of double data entry is to identify discrepancies for correction. Options include:
    a) Entering the data twice into two separate files and then comparing the two files for differences.
    b) Prepare for duplicate entry. After the first file is completed, the second file is prepared based on a key field (the unique identifier) for the first file. While entering the second file, the value is checked for each field in each record against the same record of the first file. You are warned of any discordance so that you can ensure proper recording during the second entry process. This feature is available in EpiData.
- Double entry won't identify an error if the same error is made twice. The chances of this occurring may be reduced by a second data entry operator entering the duplicate data.

11.4. Data cleaning after database closure

- **Garbage in = Garbage out**
  Remember that the quality of the results you produce is directly related to the quality of your data.
- **Before analysing a set of data, it is important to check as far as possible that the data are correct.** Errors can be made at many points in the data collection process: when measurements are taken, when the data are originally recorded, when they are transcribed from the original source, or when being entered into a computer. We can't be sure about what is correct, but we can check if recorded values are plausible. This process is called data cleaning.
- **Fixing errors requires a data analysis program such as Stata or running queries in the database program**. After problems have been identified you should go back and compare them to the written collection forms. Some typing errors are obvious and just need to be corrected in the database. Other errors will require some interpretation. If it is not possible to decide on a meaningful response, these variables should be recorded as missing. Data can be corrected in the original database and re-exported for use in the statistical package. Any corrections made in the statistical package should be documented so that any repeated analysis will achieve the same result.
- **Checking and cleaning your data takes longer than you think**. Allow sufficient time for this stage when planning a study. Once the database has been closed, as decided by the principal investigator, any data corrections must be made in the statistical package.
- **DO NOT do any data cleaning interactively in your statistical package**. Cleaning must be documented and reproducible. Data must be cleaned via a command file (eg: a do-file in Stata, a syntax file in SPSS etc). Include comments in your data cleaning command file.
- **DO NOT overwrite original variables**. It is often necessary to re-code or modify original variables. It is good practice to assign the modified values to new variables and keep the original variables unchanged. The exception to this recommendation is replacing missing codes with missing values.

Prior to Database closure, cleaning might involve:

- Data checking throughout the data entry process – can minimize data entry errors due to interpretation differences between data entry personnel
- Double data entry - any discrepancies between the two copies are checked and corrected
- 10% hard copy checks – randomly select 10% of the records and two people then compare a print out of the selected electronic data with the paper collection forms

After database closure, cleaning might involve:

- Checking there are no query codes remaining
- Identifying blanks where there should be none
- Identifying implausible values
- Inconsistency checks - "Logic checks" – run queries on the data to pull out records that don't match certain rules defined by the project e.g. age criteria
- Replacing missing codes with missing values
- Attaching variable and value labels

11.4.1. Missing Data

- **Check for blanks**. It is advisable that codes for missing data are created prior to data entry. Therefore when data is entered, the only fields that should be left blank are those with a jump, e.g. If "No" skip next question. This makes it easy to spot fields that have been skipped in the data entry process.
- **If an error is found**, ideally the value should be changed to the correct value. However, if these is no record of what the value should be, the missing value code should be used, e.g. 9=missing.
- **Tell the package which values indicate missing data**. Usually this means converting the numeric missing code to the program's official 'missing value'.

11.4.2. Logical checks

- **Check for duplicate records**. Each record in a file must have a unique key. Usually this is a single variable (e.g. idno.), but it may be a combination of two or more variables (e.g. idno and visit date).
- **Check for consistency between variables**. Data values can depend on the value of another variable. For instance, in a study of survival after a kidney transplant, information on the number of previous pregnancies is relevant only for women, who should all have a non-missing value, whilst men should all have a missing value.
- If you have a set of criteria for selecting subjects for your study, check that all participants were eligible.
- **If a measure is recorded more than once at the same time point,** the repeated values should be within a reasonable range of each other. For instance, you may decide to newborn head circumference more than once to get a more accurate value. Discrepancies between the measurements should be small.

11.4.3. Checking categorical data (e.g. yes/no or mild/moderate/severe)

It is quite a simple task to check that all data values are plausible because there are a fixed number of pre-specified values. For each of the categorical variables produce frequency tables showing all the recorded values. Alternatively, if the package allows, include statements that make explicit checks on values. For instance, if gender has been coded as: 0=female, 1=male, 9=missing, then the statement would assert that gender contains only the values 0, 1 or 9. If the statement fails, you know that the variable contains at least one dubious value.

11.4.4. Checking continuous data (e.g. height or age)

For continuous data we can specify lower and upper limits on what is reasonable. Values that fall outside this range may not necessarily be wrong. All suspicious values should be checked and any errors corrected. If a value is felt to be impossible rather than just unlikely, it should be recorded as 'missing'. *Be aware that sometimes an apparently extreme value may be valid.*

- Produce summaries showing the mean, median, variance and minimum and maximum values for each continuous variable.
- As with categorical variables, you can include statements to check for values below the expected minimum and above the expected maximum values.
- Produce a dotplot to easily spot any possible errors.

11.4.5. Checking Dates

- Check all dates are within a reasonable time span. In a study where year 7 students are surveyed, the date of birth should be about 12 years prior to the survey date.
- Check dates are in the right order, eg dob < date of 1st visit< date of 2nd visit
- Ages and time intervals can be calculated via a statistical package using the relevant dates. Check that ages and time intervals lie within the expected range. eg: negative ages indicate data error

11.4.6. Longitudinal Studies

- Where the same variable is measured at several time points for each subject, it is valuable to plot each person's sequence of recorded values to ensure that they behave reasonably.
- Check that variables that shouldn't change over time are consistent.
- Only id numbers/subjects with a baseline record should have data at later time points.

11.5. SUMMARY: Steps to good data management

- Use well designed data collection forms
- Create a data dictionary to document the project's metadata (information about the data such as its meaning, relationships to other data, origin, usage, and format)
- Use a relational database where possible
- Ensure the database design will give the required outcomes with the greatest accuracy
- Carry out data cleaning before analysing the data
- Keep a record of all analyses (e.g. STATA do files)
- Archive all data at the completion of the project (both paper based and electronic)
- And budget appropriately in your project from the start for these activities, including input from a data manager

- Data quality is achieved with a meticulous, systematic and logical approach to data management.
- If not enough care, thought and time are given, problems can occur at the analysis stage. Your analysis will then be based on invalid data, leading to false results. *You need to be obsessive*.

12. Key resources

12.1. REDCap access and support

See extensive information on the Research Education Program open access website:

- **REP Seminar: "Using REDCap for data capture and management"**
  This 1hr overview reviews the basic functionality of REDCap and introduces some of its features. Access through the Past Seminars page of our website, from Section 6. Data Management and Statistics:
  https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Past-seminars

- **REDCap Workshops and resources**
  View recordings of the REDCap Basics and REDCap Intermediate workshops hosted by Research Education Program and download the accompanying resources from our "Workshops" page on the website.
  https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Workshops

- **General access and help information including other instructional videos**:
  https://www.cahs.health.wa.gov.au/Research/For-researchers/Research-Education-Program/Workshops

12.2. Important REDCap information for CAHS staff

- Updates to the REDCap licensing terms and conditions mean a Telethon Kids Institute employee must be actively engaged in all projects using the Telethon Kids instance of REDCap.
- Projects where the entire team are external (Dept of Health employees without a Telethon Kids appointment) cannot reside on the Telethon Kids instance of REDCap.
- In the short-mid term, all existing projects set-up on the Telethon Kids instance of REDCap can remain.
- Access to a Dept of Health instance of REDCap is now available for all WA Health employees (with an active HE number and WA Health email address). See attachment below.
- REDCap support is still available to all CAHS Dept of Health based researchers through the Telethon Kids Biometrics team.
- For projects utilising the Dept of Health instance of REDCap, workshops and support/advice is still available, but there are limitations on the 'hands on' account related activities able to be performed.
- For projects with active Telethon Kids collaborations, there are no changes to the ability to use the Telethon Kids Instance or the processes by which you do this.

12.3. More useful websites

Other resources are embedded above through this document

- NHMRC: Australian Code for the Responsible Conduct of Research (2018) https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018
- NHMRC Competencies for Australia Academic Clinical Trialists (May 2018) https://www.nhmrc.gov.au/about-us/publications/competencies-australian-academic-clinical-trialists
- National Statement on Ethical Conduct in Human Research (2007) https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018
- EpiData Software link http://www.epidata.dk/
- UK Data Archives http://www.data-archive.ac.uk/
- University of Western Australia - Research Data Management Toolkit https://guides.library.uwa.edu.au/RDMtoolkit/support
- University of Western Australia - Code of Conduct for the Responsible Practice of Research. Section 2. http://www.governance.uwa.edu.au/procedures/policies/policies-and-procedures?method=document&id=UP12/25

12.4. Data Linkage Branch Training for linked data

- Free workshops for researchers and other applicants interested in applying for linked data.
- Contact: DataServices@health.wa.gov.au
- Workshops generally cover core essentials:
  - The data linkage process
  - The preparation of data
  - The datasets available to researchers
  - Ethical considerations
  - The application process

12.5. Data Manager Support

Data managers can provide advice and/or assistance across a wide range of issues such as data base set up, data entry and cleaning. It is wise to allow in your project budget sufficient funds to seek assistance from a data manager from the earliest possible stages. Collaboration may be key to achieving access to what is sometimes a limited resource. Many research institutes, universities and some hospital departments have a data manager on site available for support.

Government of Western Australia
Child and Adolescent Health Service

CAHS Research Education Program

Research Skills Seminars Series 2021

CONSUMER & COMMUNITY INVOLVEMENT

30th July 2021 | 12:30pm – 1:30pm | Perth Children's Hospital

Every researcher should be actively involving consumer or community members to improve quality and increase impact of their research. Community involvement is increasingly a requirement for funding agencies. This seminar provides a practical introduction and will cover basic principles of consumer and community involvement, the benefits and barriers, and what to put in place to get started.

Perth Children's Hospital
Level 5, 15 Hospital Ave Nedlands
Accessible via pink or yellow lifts

- OR -

Access online via Avaya Workplace

- OR -

Watch live from a hosted video-conferencing site at

- Fiona Stanley Hospital
- Lions Eye Institute
- Royal Perth Hospital

About the Presenter

Anne McKenzie AM

Anne McKenzie AM is the Community Engagement Manager at Telethon Kids Institute. She was the former Head of the Consumer and Community Health Research Network which is an enabling platform of the WA Health Translation Network.

Anne is a senior consumer representative for state and national health committees and former Chair of the Health Consumers Council WA. In 1994 Anne established the inaugural role of the Parent Advocate at PMH and has a long history of consumer advocacy with WA health services.

Click here to register online
or visit
https://20210730.eventbrite.com.au

Discover

To watch past seminar recordings, download presentation material or subscribe to our event notification newsletter, visit:
cahs.health.wa.gov.au/ResearchEducationProgram

Contact

Phone (08) 6456 0514
Email researcheducationprogram@health.wa.gov.au
Intranet cahs-healthpoint.hdwa.health.wa.gov.au

The Research Skills Seminar Series is part of the Research Education Program presented by the Child and Adolescent Health Service, Department of Research. Seminars are hosted by WA Department of Health.

Government of Western Australia
Child and Adolescent Health Service

CAHS Research Education Program

Research Skills Seminars Series 2021

KNOWLEDGE TRANSLATION

***NEW DATE* Friday 6th August | 12:30 – 1:30pm | Perth Children's Hospital**

Ensuring that research findings are translated into practice involves a systematic approach from the beginning when you are designing your research. Implementation science bridges the gap between developing and evaluating effective interventions and implementation and de-implementation in routine practice. This seminar covers key elements of implementation research; theoretical approaches, research designs, involvement of stakeholders, behaviour change interventions.

Perth Children's Hospital
Level 5, 15 Hospital Ave Nedlands
Accessible via pink or yellow lifts

- OR -

Access online via Avaya Workplace

- OR -

Watch live from a hosted video-conferencing site at

- Fiona Stanley Hospital
- Lions Eye Institute
- Royal Perth Hospital

About the Presenter

Dr Fenella Gill

Fenella is Associate Professor - Acute Paediatric Nursing at School of Nursing, Midwifery and Paramedicine, Curtin University and Perth Children's Hospital (PCH), Child and Adolescent Health Service. She leads research focused on paediatric inpatient and family experiences, safety and outcomes.

Fenella's PhD work resulted in national practice standards for critical care nurse education, incorporating the views of health consumers. Fenella holds an inaugural West Australian Health Translation Network (WAHTN) and Curtin University 2019 Early Career Fellowship in Research Translation and in 2016 Fenella was honoured as a life member of the Australian College of Critical Care Nurses (ACCCN).

Click here to register your participation

Discover

To watch past seminar recordings, download presentation material or subscribe to our event notification newsletter, visit:
cahs.health.wa.gov.au/ResearchEducationProgram

Contact

Phone (08) 6456 0514
Email researcheducationprogram@health.wa.gov.au
Intranet cahs-healthpoint.hdwa.health.wa.gov.au

The Research Skills Seminar Series is part of the Research Education Program presented by the Child and Adolescent Health Service, Department of Research. Seminars are hosted by WA Department of Health.

CAHS Research Education Program

Research Skills Seminar Series 2021

A free, open-access resource designed to upskill busy clinical staff and students and improve research quality and impact.

2021 Seminar Schedule

Updated 29 June 2021

Date	Topic	Presenter
Feb 5	**Research Fundamentals**	A/Prof Sue Skull
Feb 19	**Scientific Writing**	A/Prof Sue Skull
Mar 12	**Introduction to Good Clinical Practice**	Natalie Barber
Mar 19	**Research Governance**	A/Prof Sunalene Devadason
Apr 30	**Using Social Media in Research**	Dr Kenneth Lee
May 7	**Using REDCap for Data Capture and Management**	Telethon Kids Biometrics Team
May 14	**Survey Design and Techniques**	A/Prof Sue Skull
May 28	**Getting the most out of Research Supervision**	Prof Jonathan Carapetis AM
Jun 18	**Introductory Biostatistics**	Dr Julie Marsh
Jun 25	**Sample Size Calculations**	Dr Julie Marsh
Jul 23	**Data Collection and Management**	A/Prof Sue Skull
Jul 30	**Consumer and Community Involvement**	Anne McKenzie AM
Aug 6	**Knowledge Translation**	Dr Fenella Gill
Aug 13	**Media and Communications in Research**	Elizabeth Chester
Aug 27	**Oral Presentation of Research Results**	A/Prof Sue Skull
Sep 10	**Conducting Systematic Reviews**	Prof Sonya Girdler
Sep 17	**Involving the Aboriginal Community in Research**	Glenn Pearson & Sue Skull
Oct 22	**Rapid Critical Appraisal of Scientific Literature**	A/Prof Sue Skull
Oct 29	**Statistical Tips for Interpreting Scientific Claims**	Dr Julie Marsh
Nov 5	**Grant Applications and Finding Funding**	Tegan McNab & Sue Skull
Nov 12	**Research Impact**	Tara McLaren
Nov 19	**Ethics Processes for Clinical Research in WA**	A/Prof Sue Skull
Nov 26	**Qualitative Research Methods**	Dr Shirley McGough
Dec 3	**Innovation and Commercialisation**	REP & Telethon Kids Institute

REGISTER ➲ Follow our Eventbrite page to register throughout the year or click the hyperlinked titles

ACCESS ➲ View recordings from previous seminars

SUBSCRIBE ➲ Subscribe to receive event invitations

☎ (08) 64564585 ✉ ResearchEducationProgram@health.wa.gov.au cahs.health.wa.gov.au/ResearchEducationProgram

All seminars are held from 12:30-1:30pm in the Auditorium on Level 5 at Perth Children's Hospital and topics may be subject to change – email notice will be provided. All corresponding handouts are regularly revised and updated with attendance certificates available upon request.

CAHS Research Education Program

Research Skills Seminar Series

A free, open-access resource designed to upskill busy clinical staff and students and improve research quality and impact.

Data Collection and Management

Thank you for your interest in this seminar

Please complete this 1-minute evaluation.
Your feedback will help guide future presentations and educational activities.

How did you attend the seminar?

- ○ Live seminar at Perth Children's Hospital
- ○ Hosted video-conference on-site (e.g. FSH, Lions Eye, RPH etc.)
- ○ Online via Scopia
- ○ Viewed online recording

Please rate your agreement with the following statements:

	N/A	Strongly Disagree	Disagree	Neither	Agree	Strongly Agree
The aims and objectives were clear	○	○	○	○	○	○
The session was well structured	○	○	○	○	○	○
Presentation style retained my interest	○	○	○	○	○	○
The speaker communicated clearly	○	○	○	○	○	○
The material extended my knowledge	○	○	○	○	○	○
The additional resources were helpful	○	○	○	○	○	○

What were the best aspects of the seminar?

What changes or improvements would you suggest?

How did you hear about the seminar?
(you can select multiple answer)

- ☐ Email invitation from Research Education Program
- ☐ CAHS Newsletters e.g. The Headlines, The View, CAHS Research Newsletter
- ☐ "Health Happenings" E-News
- ☐ Healthpoint Intranet Upcoming Events
- ☐ Collegiate lounge screen or other posted promotional material
- ☐ Telethon Kids Institute screen or other posted promotional material
- ☐ Telethon Kids Institute Newsletter
- ☐ Other

Thank you!

https://cahs.health.wa.gov.au/ResearchEducationProgram