

Statistician for Starship Child Health

First name: Dug Yeo (two words)

DugYeoHan@adhb.govt.nz

Dug Yeo Han, PhD

Clinical Research Biostatistician

Research & Innovation Starship Child Health

Mob: **021 0237 1867**

Auckland District Health Board | Level 5 | ACS House | 3 Ferncroft Street |
Grafton | Auckland 1010

Website: <https://www.starship.org.nz/health-professionals/research-at-starship/>

- Statistician contribution
 - Check definition of variables
 - Check key variables are being measured and recorded
 - Check that the real question involving primary outcome can be answered.
- PI contribution – Right question for the right data
- Lead to
 - Analysis plan

Organising Data in Excel

*Organize the data as a **single rectangle***

- With **subjects as rows** and **variables as columns**
- With a **single header** row

Organising Data in Excel

Be Consistent

- Use consistent codes for categorical variables,
e.g. 'F' ≠ 'Female' ≠ 'female' ≠ 'f'
- Use a consistent fixed code or leave it as **blank** for missing values
- Use consistent subject ID, e.g. NHI9999
 - Give another variable name for multiple admission
- Use consistent format for all dates, e.g. DD/MM/YYYY or DD-MM-YYYY
- Be careful about extra spaces within cells (spaces at the beginning and/or end), e.g. 'NHI9999' ≠ ' NHI9999 ' ≠ 'NHI9999 ' ≠ ' NHI9999 '

Organising Data in Excel

Choose informative names for variables in column headers

- Record units in column headers, e.g. Age_yr
- Use unique name
 - E.g. Weight measures at admission and at discharge then and named as Wt_adm and Wt_dis
- Do not use
 - Spaces
 - Punctuation
 - Special characters
 - Except underscore

Organising Data in Excel

Example for variable names

In Excel

CaseNo	Hospital	Age_yr	Gender	Ethnicity1	Ethnicity2	Ethnicity3	NZDep2013	Liver_disease	Renal_disease	Chronic_disease	Longterm_medication	Shortterm_medication
ID001	Christchurch	14y 6m (14.5)	F	Samoan			3	No	No	No	No	

Example

Not-So-Good way to set-up	Better Way
Case No.	CaseNo
Hospital 1	Hospital
(Formula) Age at time of report	Age_yr
Q2 DDB: Gender	Gender
Q4 DDB Ethnicity Tick all that apply(1/9 options)	Ethnicity1
Q4 DDB Ethnicity (2/9 options)	Ethnicity2
Q4 FT: Other Ethnicity (3/9 options)	Ethnicity3
Postcode_average_NZDep_2013	NZDep2013
Q12 DDB Did the child have past medical history liver disease	Liver_disease
Q12 DDB Did the child have past medical history renal disease	Renal_disease
Q12. Free Text Did the child have past medical history other chronic disease (excluding mental health) free text	Chronic_disease
Q13. DDB At time of ingestion was the child receiving any other long term medication	Longterm_medication
Q13. DDB At time of ingestion was the child receiving any other short term medication	Shortterm_medication

Organising Data in Excel

Create a data dictionary

Contains:

- The exact variable name as in the data file
- An explanation of what the variable means
- The measurement units
- Expected range of values *e.g. dose of medication*
 - Function called 'Data Validation' in Excel is useful

Organising Data in Excel

Do not include calculations in the raw data files

- Panko (2008) reported that in 13 audits of real-world spreadsheets, an average of 88% contained errors
- Let statistician do the calculations

Do not put continuous measurements into categories when you enter the data

- Let statistician do the categorisation

Each cell in your cases x variables spreadsheet should contain *only one piece of data*. Don't try to put in more data by *colour coding* the cell (or *changing the font*).

Do not put comments in cells

What to think about *before* you start your trial

