

NatCen

Social Research that works for society

Deriving variables: A Data Managers' guide to DV best practice

Jess Bailey,

The National Centre for Social Research

14th February 2020



Importance of reproducibility

■ Health Survey for England

- Cross sectional survey
- Run every year since 1991 (carried out by UCL and NatCen since 1994)
- 650-800 DVs created per survey: 600 repeated; 50-150 new

■ British Social Attitudes

- Cross sectional survey
- Run (almost) every year since 1983
- 500 DVs created per survey: 450 repeated; 50 new

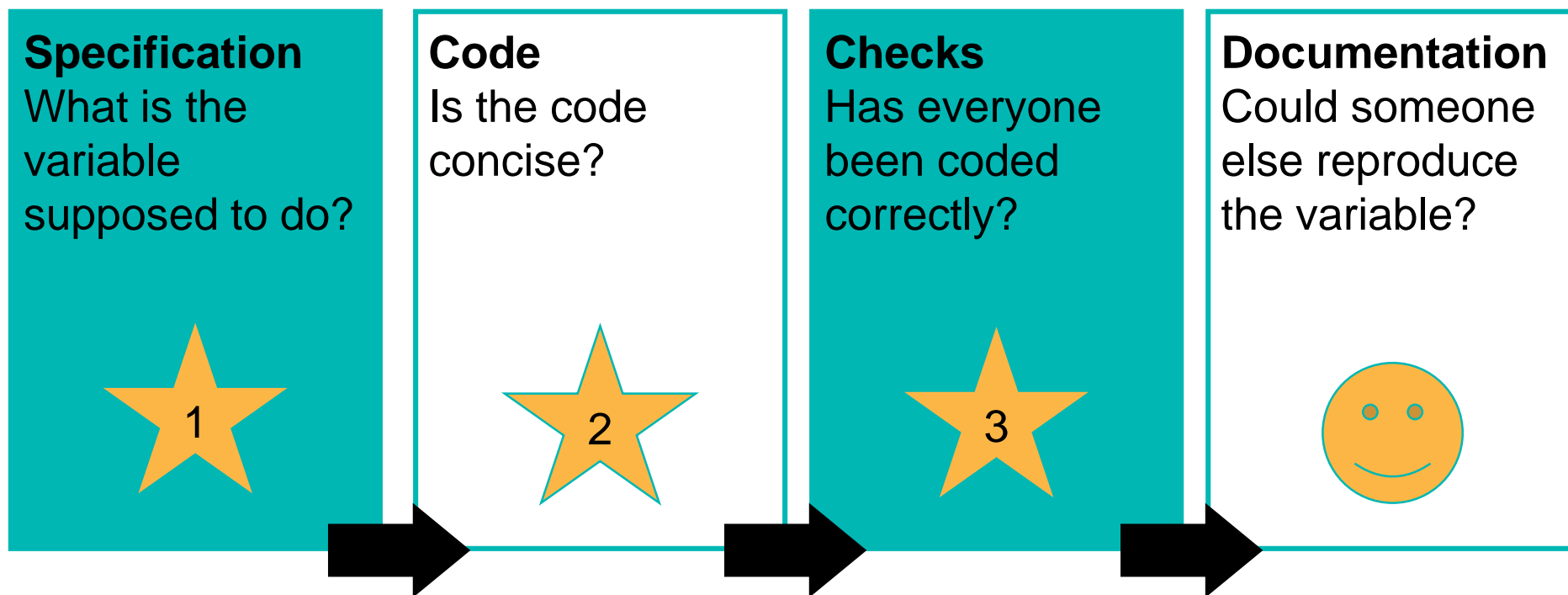
■ NatCen Panel

- Longitudinal panel survey
- Run every other month since May 2016
- 250-300 DVs created per survey: 150 repeated each month; 100-150 new



How to achieve DV excellence?

Comprehensive and clear...



Comprehensive Specification

Specifying DVs

1. Specify variable name and variable label, value codes and labels
2. A short description of what the variable is supposed to do

3. Input variables used for the derivation
4. How to handle missing values e.g. don't know, refusal, system missing/blank

Comprehensive Code: Part 1

Curating DVs

Catalogue by type/module

Consider order of derivations if there are dependencies

Set the metadata upfront

Variable name: Check duplicate names

Variable and value labels: present during computation



Comprehensive Code: Part 2

Coding DVs

Missing values off

Compute all cases to non standard code, e.g. -77

Be *explicit*, e.g. avoid: NE, ELSE=COPY

Look for *efficiencies*; i.e., least lines possible: RECODE, DO REPEAT



Comprehensive Checks: Part 1

Step 1 Turn on labels and codes for display – run a **frequency**

Step 2 Check against the variables used for the derivation – run **cross-tabulations** or **c-tables**

Comprehensive Checks: Part 2

Step 3 Retain but comment out frequencies and cross-tabulations. A second person should cross check the DVs.

Step 4 Specify a check to produce an error report when cases are missing e.g., equal to -77

Comprehensive Documentation

Archiving data

- Data sets
- DV syntax
- Data documentation
- Questionnaire specification
- User Guide



Contact

Jess Bailey

Data Management Director

T. 020 549 7015

E. jess.bailey@natcen.ac.uk

Visit us online, natcen.org.uk

NatCen

Social Research that works for society