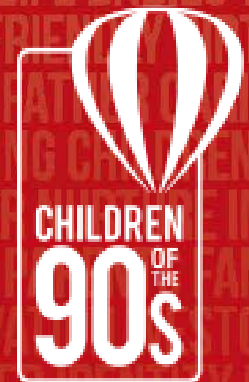


# Disclosure control in ALSPAC

~~Olly Butters~~ – Senior data manager  
Paul Burton

18/1/17



Avon Longitudinal Study  
of Parents and Children

Supported by  
**wellcome**trust



# Overview

- Summary info
- Unique collaborator ID for data requests
- Triplets & Quadruplets
- Text data

# Bespoke data extracts

- Each data extract is built with just the requested variables.
- All IDs are replaced with an ID unique to this data request.
- Stops the aggregation of variables across projects.



## Summary info in data dictionary

- Data dictionary has frequencies for all questions and clinic measures.
- Many have a frequency of 1.
- Where a question is deemed identifiable for small numbers the bands are aggregated together (e.g. age of participant when data collected).
- Based on experience/judgement of person building the data set, not a rigid set of rules.

## Triplets and quadruplets in the data

- Core built data set holds triplets and quadruplets.
- Last step before giving bespoke data set to collaborators is to take triplets and quadruplets out entirely.
- Manually done each time.

## Text data

- Questions range from expected single word answers (e.g. medications) to long free text.
- All answers are manually checked for disclosure - e.g. looking for names etc.
- Very time consuming.

# Geographical data

- Hold addresses, locations of schools, places of work etc.
- Not allowed out of the building - researchers come in to use it.
- Any output looked at before being allowed to leave building.