

Overview over traditional disclosure avoidance methods

Ian Schmutte and Lars Vilhuber

Purpose of disclosure avoidance methods

Discuss legal and ethical requirements. Define terms, with reference to external sources (de-identification, anonymization, pseudonymization, European, US, and Indian legislation). Link to "safe data" and "safe outputs". Reference glossaries where available.

(1-1.5 pages)

Methods

Define the methods used for disclosure avoidance. (removal, coarsening of various kinds, swapping, data synthesis, formal privacy). Should reference WP 22, **accessible** guidelines by various entities including J-PAL, FCSM. Note: WP-22 is being revised, and will be released AFTER the handbook is released - so needs to have some forward looking elements. Reference glossaries where available.

(2-3 pages)

Metrics

How do you measure risk, and the reduction in risk achieved by applying above methods? Mention uniqueness criteria, k-anonymity, l-diversity, matching metrics, etc.

(2-3 pages)

Tools

A sketch with lots of links to Stata and R packages that implement disclosure avoidance methods. Some simple methods to compute metrics as well (usually integrated).

- Summary of methods/packages in Stata
- Summary of methods/packages in R

(1-2 pages)

Impact on analyses

What is the impact on analyses when using protected data (reprise Abowd-Schmutte, BPEA), but also Chetty + Friedman (AEA P&P, JPC) and Green-McKinney-Abowd-Vilhuber (increased variance) as well as the various papers on MOE/ACS. Consider Dwork on alignment of DP methods and safe inference.

- Reference standard methods that are related
 - Moulton correction (data aggregated - for whatever reason)
 - Using ACS MOE as an example of incorporating uncertainty (regardless of source)
- Parts of Abowd-Schmutte, BPEA

(1-2 pages)

Biblio databases on Zotero

- IDEA: 2400539
- 2454585 privacy__bibliography