Balancing Privacy and Information Disclosure in Interactive Record Linkage with Visual Masking

Hye-Chung Kum 1,3,4  Eric D. Ragan 2,3  Gurudev Ilangovan 3,4  Han Wang 3,4

1 Department of Health Policy & Management, 2 Department of Visualization, 3 Department of Computer Science & Engineering, Texas A&M University
4 Population Informatics Lab (https://pinformatics.org/)

Interactive Record Linkage

- Same person?
- Data source 1
- Data source 2

Visual Markup

- Highlight discrepancies
- Name frequency meta-data
- Empty fields
- Different characters
- Extra characters
- Transposed values
- Name or Date Swaps
- Major field differences

Data Masking

- Hide data details for privacy
- Same values
- & & Different values

Experiment

- 104 participants
- 90 minutes
- Tutorial
- Main trials (26 linkage pairs)
- Additional practice and questionnaires

Results

- No effects of visual markup
- 84.8% 84.1% 84.5% 78.1% 74.5%
- 6.4% 10% drop drop p < 0.05

Confidence

- Conclusion
- For legitimate data work such as data integration and verification using PII data, different people need to have access to personal information, which sacrifices the personal privacy of those whose data is stored.
- Often, the primary methods for handling privacy concerns are either to restrict data access at the expense of data utility, or to open the data to more people to improve throughput and utility at the expense of reduced privacy.
- Our study results demonstrate that it is possible to significantly reduce PII disclosure without noticeably affecting decision accuracy with appropriate meta-data.
- Moreover, when legal requirements only allow for de-identified data access, use of well-designed interface can significantly improve data utility.

Research supported in part by the Patient Centered Outcomes Research Institute ME-1602-34486.