

International Journal of Population Data Science

Journal Website: www.ijpds.org



Designing An Ambiguity-Aware Clerical Review Strategy

Joseph Lam¹, Amaia Imaz Blanco², Efrosini Setakis², Jonathan Laidler², Jonathan Pearson², Katie Harron³, Mario Cortina-Borja³, Peter Christen⁴, and Giulia Mantovani²

¹NHS England, London, United Kingdom; UCL Great Ormond Street Institute of Child Health, London, United Kingdom

²NHS England, London, United Kingdom

³UCL Great Ormond Street Institute of Child Health, London, United Kingdom

⁴School of Computing, The Australian National University, Canberra, Australia; Scottish Centre for Administrative Data Research, University of Edinburgh, Edinburgh, United Kingdom

Clerical review of candidate record pairs remains the de-facto gold standard for evaluating record linkage, but is resource-intensive and often designed informally. We propose a design-based framework that treats clerical review as finite-population sampling over strata defined by match weight, identifier agreement pattern, record-level ambiguity and demographic group. Bands are constructed from deciles of model-based match probability; within bands, strata are formed by agreement patterns and record ambiguity. A band-wise margin-of-error profile encodes substantive priorities (e.g. tighter targets in high-score bands), while a single global scaling parameter enforces an overall clerical budget. From deduplicating a labelled training dataset using Splink (50,000 records; ~478,000 candidate pairs), we compared performance of baseline design reviewing ~23% pairs with “budget” design reviewing ~7% pairs. Under the baseline design, estimates of global (4.2%) and

band-specific match rates are accurate and the sampled distributions of agreement patterns, demographics (such as gender) and ambiguity track the population reasonably well. Under the “budget design”, global linkage error estimation roughly doubles (9.3%), band-level errors increase most in mid-score bands where matches, non-matches and ambiguous cases are intermixed, but accuracy in the highest bands and the band-wise record ambiguity are largely preserved. We focused clerical effort on records with most ambiguous candidate pairs, at the cost of reduced representativeness in agreement patterns and gender. The framework generalises to other clerical review goals and can incorporate gold standards as priors for calibration. It makes the trade-offs between clerical workload, precision, representativeness and ambiguity coverage explicit and therefore negotiable.

