

Supplementary Table 1. Definitions of linkage terms.

Linkage Term	Definition
Data elements	Variables/fields containing identifying information that can be used to determine if records belong to the same individual (e.g., first name, last name, ZIP code of residence).
Discriminating power	For a given data element, the probability that values are the same for a pair of records even though the records do not belong to the same individual. The discriminating power approximates the probability of agreeing by chance.
Hierarchical deterministic linkage	In a deterministic linkage, records are joined based on having identical values for one or more data elements (e.g., full date of birth) or parts of those elements (e.g., month and year of birth). A hierarchical deterministic linkage includes a series of passes, generally with the strictest criteria first; once a match is found, the record is not processed in subsequent passes.
False match	Incorrectly matched records that in truth belong to different individuals.
False match probability	Likelihood that two matched records in truth belong to different individuals. Calculated as $1 - \text{match probability}$.
False match rate	Proportion of accepted pairs that are false matches. It is calculated as the sum of the false match probabilities among all accepted pairs (i.e., pairs with a match probability above an assigned threshold) divided by the total number of accepted pairs.
False non-match	Records that in truth belong to the same individual but were not matched during the linkage process.
False non-match proportion	Estimated proportion of records that were not matched with another record but in truth belong to the same individual.
Join (or block) criteria	Criteria used to bring together records for additional evaluation. Records that have exactly the same values on selected criteria are further evaluated using match criteria, while records that do not have the exact same values on these criteria are not evaluated. Consequently, join criteria reduce the number of comparisons and allow for more efficient processing. Join criteria differ for each pass so that, potentially, different pairs of records are evaluated.
Match criteria	Criteria used to evaluate pairs of records and assign match probabilities. Match criteria are the same across passes and can allow for inexact values (e.g., 1 typo in Social Security Number, date of birth matches +/- 3 days).
Match probability	Likelihood that two records in truth belong to the same individual. These estimates consider discriminating power of data elements included in the algorithm as well as their reliability.
Passes	Iterations of a linkage algorithm that use different criteria to join records for evaluation.
Probabilistic linkage	Linkage in which compared records are assigned a probability (i.e., likelihood) that they belong to the same individual. Probabilities are based on agreement and disagreement of values for multiple data elements. Records with a probability above an assigned threshold are accepted and records below that threshold are rejected.
Reliability	For a given data element, the probability that values are the same for a pair of records that in truth belong to the same individual. This factors in how error-prone a variable is and approximates $1 - \text{error rate of the variable}$.
Single file match	Linkage process that involves only one data set. Within that single data set, all of the records that meet join criteria are compared, one pair of records at a time. This contrasts with two file matches, in which records from data set A are compared against records from data set B (and not to other records in data set A).
Set	All of the records deemed to belong to the same individual as a result of the linkage process.
True match	A correct identification that records in truth belong to the same individual or that records in a set in truth belong to one individual.
True match proportion	Proportion of matched records that are true matches. Or, the proportion of sets for whom all of records within the set in truth belong to one individual.