

Appendix A

Record Matching Algorithms

Record matching algorithms are used to perform front-end record searches, to link records across disparate systems, and/or to identify possible duplicate records on the back end. Algorithms are defined as follows.

- **Basic algorithm:** Simplest technique for matching records; utilizes deterministic matching.
 - The data elements must match exactly (or have an exact partial match) in order to return a particular record match.
 - Comparisons are usually made only on name, birth date, Social Security number (SSN), and sometimes gender.
- **Intermediate algorithm:** Uses more advanced programmatic techniques than basic algorithms to compare records.
 - Phonetic encoding systems and sometimes equivalent name tables are utilized to counter misspelled names and nicknames. Arbitrary/subjective field match weights are assigned to key patient-identifying attributes such as last name, first name, date of birth, and SSN, resulting in a record match weight score.
 - May utilize programs to address transpositions, digit rotations, and typographical errors.
- **Advanced algorithm:** Utilizes the most sophisticated tools for matching records and relies on mathematical and statistical theory.
 - Core intelligence includes probabilistic theory and mathematical/statistical models, which are applied to determine likelihood of a match on specified data elements.
 - It includes machine learning and neural networks, which use forms of artificial intelligence that simulate human problem solving.