Linkage of Health Information in the Irish Health System: Challenges & Solutions

Kevin Patrick O'Brien

Medical Data is expected to double every 73 days by 2020
(A lot of clinical measurements)

Hospital Data

- Hospital In-Patient Enquiry (HIPE) system.
- Computer based system for administrative health data.
- Systematically collecting data since early 90s.
- Formerly part of ESRI, now part of HSE.

Children's Database

What is childrensdatabase.ie?

Childrensdatabase.ie was created to provide access to research and information on children for policy-makers, Government departments, academics, voluntary organisations and the general public. This website is maintained by the Department of Children and Youth Affairs (DCYA).

What is available on childrensdatabase.ie?

- Irish Government policy documents
- Annotated websites

Annotated websites

A collection of databases, specialist libraries, Irish, European and international resources, organised as follows:

- **Agencies and organisations**
- **Databases**
- **Data sources**

THE IRISH PRIMARY CARE RESEARCH NETWORK

The Irish Primary Care
Research Network

Pooling Data

- Health Boards and GPs need to be brought on board.
- Communicate benefits of sharing and pooling data for syudying population health.
- Negative cases used in contrast with positive cases.
- Privacy Issues, Statistical Disclosure Control.

NHS details released against patients' wishes, admits data body

Health and Social Care Information Centre failed to log requests of up to 700,000 patients not to pass on details and says issue 'may take some time' to resolve

The Health and Social Care Information Centre has said technical issues led to the data being shared. 'Basically it's a mess,' said Dr Beth McCarron-Nash of the General Practitioners Committee. Photograph: Dominic Lipinski/PA

Unique Idenifiers

- In Database Science - Primary Keys
- Which Kevin O'Brien are we talking about?
- PPS number.
- Which Crescent Court are we talking about? (Later)

Personal Identifers

- *Personnumber*
- Unique Identifier for all residents of Sweden

890102-3286

Födelsedatum. Två siffror för respektive år, månad och dag   Födelseplats   Kön   Kontrollsiffra

Wikipedia.com

Unique Identifiers for Medical Records

- Various Regional Healthboards.
- Public and Private hospitals.
- Can't assume same operating procedures for medical data.
- Is the available data machine readable?
- (Open Data 5 star system).
- Manual Data Cleaning : deleting meta data.

Probabilistic Record Linkage

Probabilistic record linkage.... takes a different approach to the record linkage problem by taking into account a wider range of potential identifiers, computing weights for each identifier based on its estimated ability to correctly identify a match or a non-match, and using these weights to calculate the probability that two given records refer to the same entity. (Wikipedia)

Record Linkage - R Packages

Statistical Matching and Record Linkage

- Package [StatMatch](#) provides functions to perform statistical matching between two da number of common variables. It creates a synthetic data set after matching of two data aproach or via hot-deck.
- Package [RecordLinkage](#) provides functions for linking and deduplicating data sets.
- Package [MatchIt](#) allows nearest neighbor matching, exact matching, optimal matching amonst other matching methods. If two data sets have to be matched, the data must co including a factor variable which includes information about the membership of each
- Package [stringdist](#) can calculate various string distances based on edits (damerau-lev levenshtein, optimal sting alignment), qgrams (q-gram, cosine, jaccard distance) or he winkler).

Record Linkage - R Packages

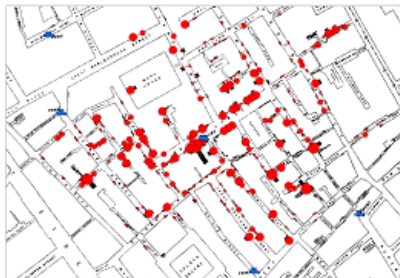# The RecordLinkage Package: Detecting Errors in Data

by Murat Sariyar and Andreas Borg

Abstract Record linkage deals with detecting homonyms and mainly synonyms in data. The package **RecordLinkage** provides means to perform and evaluate different record linkage methods. A stochastic framework is implemented which calculates weights through an EM algorithm. The determination of the necessary

age to a classification problem.

The package **RecordLinkage** is designed to facilitate the application of record linkage in R. The idea for this package evolved whilst using R for record linkage of data stemming from a German cancer registry. An evaluation of different methods thereafter lead to a large number of functions and data structures. The organisation of these functions and data structures as an R package eases the evaluation of

Spatial Analyses

Source: *eircode.ie* and *Wikipedia.com*

Unique Location Identifiers

- Population Health Analyses.
- John Snow's Cholera Map.
- How would we do something like that now?
- Eircode (proprietary system).
- Can locations be identified as "near" based on codes?