

## A national infrastructure for linkage of biobanks to medical and socioeconomic registries

G. van Grootenhuis<sup>1</sup>, J.H. Smit<sup>1</sup>, O.H. Klungel<sup>2</sup>, W.J. de Bruijn<sup>2</sup>, B.F.M. Bakker<sup>3</sup>, D.I. Boomsma<sup>4</sup>, J.A. Bovenberg<sup>5</sup>, F.E. van Leeuwen<sup>6</sup>, C.T.M. Thijss<sup>7</sup>.

<sup>1</sup> Department of Psychiatry, VU University Medical Center / GGZ inGeest, Amsterdam; <sup>2</sup> Division of Pharmacoepidemiology & Clinical Pharmacology, Utrecht University; <sup>3</sup> Statistics Netherlands, Den Haag / VU University Amsterdam; <sup>4</sup> Department of Biological Psychology, VU University Amsterdam; <sup>5</sup> Legal Pathways, Ardenhout; <sup>6</sup> Department of Epidemiology, The Netherlands Cancer Institute, Amsterdam; <sup>7</sup> Department of Epidemiology, Maastricht University Medical Centre.

## Background

Record linkage is the combination of information from different data sources, at the level of an individual.

Record linkage is a highly efficient approach to enrich existing biobanks or databanks containing socio-economic, clinical, or other (research) data. This approach can serve as an alternative to the costly collection of additional data or biomaterials within a scientific study. Additionally, it stimulates the (re)use of previously collected data and biomaterials.

The challenge in such linkages is to correctly match records from different sources when a unique identifier is not available. An important aspect is that the privacy of the subjects should not be violated. In this project, both the methodological and legal aspects of record linkage are investigated.

## Biolink NL

In the Netherlands, many high-quality medical and socioeconomic registries are available that are of interest for linkage with biobanks. Biolink NL aims to improve the efficiency and quality of linkage of biobanks to medical and socioeconomic registries, in conformity with statutory and consent obligations to participants. Ultimately, we will publish best practice guidelines for data sharing and record linkage.

This three-year project will be finalized in December 2014 and focuses on the following topics:

- a) Inventory of Dutch biobanks and building a searchable catalogue
- b) Methodology, such as data preparation and linkage algorithms
- c) Demonstration and validation of linkage of specific data sources
- d) Ethical, legal and social issues.

## Landscaping and catalogue

We are building a catalogue in which basic information about the contents of Dutch biobanks can be found. At a later stage, overlap between data sources can be determined. The infrastructure can be used as an internal tool for data managers to select overlapping research populations.

## Concepts in record linkage

**Deterministic linkage:** exact matching of variables

**Probabilistic linkage:** partial matching of variables (errors tolerated)

**Precision:** an indicator of the correct vs. the incorrect matches

**Sensitivity:** an indicator of the found vs. the missed matches

**Pseudonymization:** (irreversible) encryption of personal identifiers

**Trusted third party:** an external office that processes linkage variables

## Simulation study

When linking existing databanks, evaluation of the results is problematic because the true matches between data sources are unknown. For this reason, a simulation study was carried out in which different linkage approaches were compared. The simulated datasets were designed to resemble real databanks.

## Demonstration projects

Three demonstration projects were selected, each with their own specific attributes when it comes to population characteristics, number of subjects, and expected overlap of subjects between two sources. At a later stage, additional linkages may be added.

### Nederlandse Kanker Registratie – OMEGA cohort

**Research question:** can record linkage provide information about the incidence of (breast) cancer after an IVF treatment?

**Challenge:** OMEGA records have been collected since 1983. Linkage variables may have changed in the meantime.

### Nederlands Tweelingen Register – Achmea Health Database

**Research question:** can record linkage be used to confirm and enrich health care data collected through questionnaires?

**Challenge:** young twins are difficult to distinguish, as most of the linkage variables (except first name and sex) are identical.

### Stichting Farmaceutische Kengetallen – KOALA cohort

**Research question:** can record linkage be used to confirm and enrich pharmacological data collected through questionnaires?

**Challenge:** SFK is a vast database, but linkage variables are stored decentralized at individual pharmacies.

## Ethical, legal and social issues

Ethical, legal and social issues (ELSI) are crucial elements in the development of best practice guidelines for data sharing and record linkage. A wiki website was made which covers topics such as privacy, data ownership, informed consent and governance in the Netherlands:

[www.bbmri-wp4.eu/wiki/index.php/Netherlands](http://www.bbmri-wp4.eu/wiki/index.php/Netherlands)



## Contact

Gerard van Grootenhuis

Phone: 020 - 788 5678

Email: g.vangrootenhuis@ggzingeest.nl

Willem de Bruijn

Phone: 088 - 755 0495

Email: w.debruijn-2@umcutrecht.nl

[www.biolink-nl.eu](http://www.biolink-nl.eu)

