International Journal of Population Data Science

Journal Website: www.ijpds.org

Linkage of health insurance data to the Netherlands Twin Register

van Grootheest, Gerard[1*], Ariel, Adelaide[1], Glasner, Tina[2], Verkerk, Bep[1], van Beijsterveldt, Toos[2], van der Laan, Jan[3], de Groot, Mark[4], Visser, Sipke[5], Bakker, Bart[2,3], and Boomsma, Dorret[2]

[1]GGZ inGeest
[2]VU University Amsterdam
[3]Statistics Netherlands
[4]Utrecht University
[5]Mondriaan Foundation

Objectives

We compared four different approaches to record linkage, demonstrating its potential in enriching the Netherlands Twin Register (NTR) with healthcare data from a large health insurance provider (AHD), which covers approximately 26% of the population. Challenges included the fact that young twins share most of their linkage variables and overlapping identification numbers were absent.
We extensively validated the linkage results and complemented them with an indicator of representativeness.

Approach

Subjects born since 1986 were selected in the NTR (N=30,383) and AHD (N=1,532,675). Linkage variables were cleaned and harmonised and linkage keys were chosen that were strong enough to uniquely identify most subjects in either dataset. The four linkages were deterministic, probabilistic, probabilistic with Jaro-Winkler distance, and probabilistic using encrypted identifiers.
Validation took place by reviewing linkage variables and information like family relations; linked record pairs were categorised into correct, possible or false links. Furthermore, the consistency of linked data was reviewed focusing on Attention Deficit Hyperactivity Disorder (ADHD), a prevalent condition among children. Finally, we quantified the impact of the linkage on representativeness of the target population.

Results

Probabilistic linkage with distance calculation linked 7944 NTR subjects to the AHD with 82% correct, 17% possible, and 1% false links. It encompassed nearly all links identified by any of the other methods, which were more conservative.
Information about medication was available in the NTR dataset for 40% of the linked subjects. Validation of the information about ADHD inside linked records revealed few discrepancies between the two datasets. Linkage had a modest effect on the representativeness of the population. The current linkage has confirmed a number of known ADHD cases and identified 146 new subjects using methylphenidate and 101 receiving psychiatric care for ADHD, who were not previously recognized as ADHD cases within the NTR.

Conclusion

Record linkage with sources of data such as a health insurance database can be an efficient way of data collection in cohort research. Our study shows that record linkage is also possible for twin pairs. Although a minority of all NTR subjects were covered by the AHD, the linkage resulted in an enrichment of the NTR dataset. Exact information about the dosage and frequency of drugs was obtained without contacting subjects with detailed questionnaires. With a total of almost eight thousand retrieved records, the size of the linked dataset is sufficiently large for epidemiological research of non-rare conditions.

*Corresponding Author:
Email Address: g.grootheest@ggzingeest.nl (G. van Grootheest)

http://dx.doi.org/10.23889/ijpds.v1i1.45