Executive Summary

The EMIF-Platform is based on a federation of data sources with widely heterogeneous characteristics. The diversity of the data sources in the project represents a unique resource for epidemiological studies in Europe. However, this heterogeneity does make combining evidence from the different sources and interpretation of results an ambitious challenge.

The first part of this deliverable reports on the results of the initial round of primary data extraction (PDE). This PDE exercise was performed as a proof of principle of data sharing and data extraction from the EHR databases. The new Jerboa Reloaded PDE module generates several statistics on the population of the data sources, for example the amount of person time, number of subjects in a year, etc. The results can be uploaded to the EMIF Catalogue to complement the fingerprinting (questionnaire) data.

The second part of the deliverable describes a first version of a data derivation workflow that supports identification of subjects in a data source who have a specific clinical condition. Harmonization is performed both on the terminology level as well as on the extraction algorithms to enable a better comparison of the results from different databases. The workflow is supported by prototypes of tools that need to be further improved in the EMIF platform. Identification of type 2 diabetes mellitus (T2DM) was used as a test case for the data derivation workflow.

Contacts

EMIF-Platform: Johan van der Lei – Nigel Hughes
j.vanderlei@erasmusmc.nl - nhughes@its.jnj.com