

Description of the Mondriaan - Biolink-NL catalogue

Sipke Visser, Willem de Bruijn, Olaf Klungel.
Published on www.biolink-nl.eu, August 2015.



For more information, please contact prof. dr. O.H. Klungel:
Professor of Pharmacoepidemiologic Methods
Utrecht University, Faculty of Science
Utrecht Institute for Pharmaceutical Sciences (UIPS)
Division of Pharmacoepidemiology & Clinical Pharmacology
PO Box 80 082, 3508 TB Utrecht, the Netherlands
Phone +31 253 7324
E-mail: O.H.Klungel@uu.nl

Introduction

We have built a catalogue demonstrating the possibility to link different data sources together on subject (patient) level. Every patient within the catalogue is represented with a unique identifier, regardless of the number of sources in which the patient was found. Therefore it is possible to query over multiple sources.

An important characteristic of the catalogue is that the data of multiple sources are stored in a well-defined, uniform format. The catalogue contains heterogeneous data, making the catalogue very rich. However, the underlying registries decide which variables are uploaded into the catalogue; some sources provided more or more detailed variables than others. Obviously, the availability and precision of the provided variables determine the possibilities for querying the catalogue.

In this document we demonstrate a couple of catalogue usages. It is important to realize that the sources supplying the data determined which variables are available within the catalogue. Some sources decided that only a patient identifier should be available. Although detailed information about such patients may not be available in the catalogue, it is still possible to determine overlap with the other data sources.

An important thing to keep in mind is that the catalogue is not designed to be so detailed that scientific research can be performed on its content, but to enable quick feasibility studies before a complex route of data sharing is initiated.

Table 1. The sources present in the Biolink-NL catalogue in 2015.

Registry	number of subjects	Description
ABCD	12,097	paediatric cohort
Achmea	1,025,628	health insurance data
IADB	1,150,447	pharmacy database
JHN	775,968	Julius GP Network
KOALA	1,704	paediatric cohort
LifeLines	162,000	biobank
NESDA	2,367	biobank
PCR-MN	697,530	psychiatric case registry
PIAMA	3,167	paediatric cohort
Saltro	39,722	laboratory data
SFK	3,886,301	foundation for pharmaceutical statistics
ZGA	392,727	care group Almere

Demonstration of catalogue queries

Example 1.

There are three paediatric cohorts in the catalogue: ABCD, PIAMA and KOALA. Is there any overlap between the paediatric cohorts themselves? Indeed, we found one person to be present in both ABCD en KOALA.

Example 2.

How many children from the paediatric cohorts can be found in ZGA data?

paediatric cohort	ZGA	overlap
Piama	ZorgGroep Almere (ZGA)	1
ABCD	ZorgGroep Almere (ZGA)	185

There is no overlap found between KOALA and ZGA. In total data of 186 children from the paediatric cohorts can be enriched with ZGA data.

Example 3.

How many children from the paediatric cohorts can be found in other sources? Note that NESDA turned out to link to no source.

paediatric source	other source	overlap
-------------------	--------------	---------

paediatric source	other source	overlap
paediatric cohorts	Achmea	366
paediatric cohorts	JHN	84
paediatric cohorts	Lifelines	26
paediatric cohorts	PCR-MN	74
paediatric cohorts	Saltro	10
paediatric cohorts	SFK	3208
paediatric cohorts	ZGA	186

We can split the paediatric cohorts into its underlying registries.

paediatric source	other source	overlap
ABCD	Achmea	160
ABCD	JHN	59
ABCD	Lifelines	2
ABCD	PCR-MN	33
ABCD	Saltro	9
ABCD	SFK	2791
ABCD	ZGA	185
KOALA	Achmea	28
KOALA	JHN	22
KOALA	Lifelines	2
KOALA	PCR-MN	4
KOALA	Saltro	1
KOALA	SFK	281
Piama	Achmea	178
Piama	JHN	4
Piama	Lifelines	22
Piama	PCR-MN	37
Piama	SFK	137
Piama	ZGA	1

Example 4.

ZGA has provided detailed data, including a variable indicating the ICPC code. This enables answering questions like “How many children from the paediatric cohorts are diagnosed with ADHD?” We queried on ICPC code P20. In addition, we split the result by gender where 1 means male and 2 means female.

paediatric source	ZGA	gender	#children
ABCD	ZorgGroep Almere (ZGA)	1	7
ABCD	ZorgGroep Almere (ZGA)	2	2

So 9 children were found, and they all came from the ABCD cohort.

Example 5.

Instead of ADHD we can query the catalogue to find children with asthma or COPD, represented by ICPC coded R96 and R95, respectively. Here we show the full ICPC code.

paediatric source	ZGA	ICPC	#children
ABCD	ZorgGroep Almere (ZGA)	R96	13
ABCD	ZorgGroep Almere (ZGA)	R96.01	3
ABCD	ZorgGroep Almere (ZGA)	R96.02	2

Example 6.

All three paediatric registries cooperated for the catalogue and agreed to supply the same set of variables. One of those variables is the answer to the question if the children have used a bronchodilator. Therefore it is possible to compare the use of bronchodilators with the diagnosis of asthma within the ZGA data.

paediatric source	ZGA	bronchodilator	ICPC	#children
ABCD	ZorgGroep Almere (ZGA)	0	R96	17

There are no children found within ZGA who did use a bronchodilator. But we found other ICPC codes, enabling selection on other diagnoses.

paediatric source	ZGA	bronchodilator	ICPC	#children
Piama	ZorgGroep Almere (ZGA)	0	other	1
ABCD	ZorgGroep Almere (ZGA)	0	other	166
ABCD	ZorgGroep Almere (ZGA)	0	R96	17

Example 7.

Here we show other sources that could be linked together.

source 1	source 2	overlap
IADB	Achmea	2240
JHN	Achmea	364704
LifeLines	Achmea	151
PCR-MN	Achmea	188229
Saltro	Achmea	22780
SFK	Achmea	334041
ZGA	Achmea	17328
LifeLines	IADB	38349
SFK	IADB	72098
Achmea	JHN	364704
LifeLines	JHN	291
PCR-MN	JHN	115243
Saltro	JHN	20284
SFK	JHN	257652

source 1	source 2	overlap
ZGA	JHN	898
Achmea	LifeLines	151
IADB	LifeLines	38349
JHN	LifeLines	291
PCR-MN	LifeLines	147
Saltro	LifeLines	8
SFK	LifeLines	19300
ZGA	LifeLines	135
Achmea	PCR-MN	188229
JHN	PCR-MN	115243
LifeLines	PCR-MN	147
Saltro	PCR-MN	7818
SFK	PCR-MN	122554
ZGA	PCR-MN	16838
Achmea	Saltro	22780
JHN	Saltro	20284
LifeLines	Saltro	8
PCR-MN	Saltro	7818
SFK	Saltro	18705
ZGA	Saltro	54
Achmea	SFK	334041
IADB	SFK	72098
JHN	SFK	257652
LifeLines	SFK	19300
PCR-MN	SFK	122554
Saltro	SFK	18705
ZGA	SFK	10380
Achmea	ZGA	17328
JHN	ZGA	898
LifeLines	ZGA	135
PCR-MN	ZGA	16838
Saltro	ZGA	54
SFK	ZGA	10380

Example 8.

So far we have shown only two sources being linked together. Of course we can link more sources. For example, what is the overlap between Saltro, JHN and Achmea?

source 1	source 2	source 3	overlap
Saltro	JHN	Achmea	17663

What is the overlap between on one hand JHN and PCR-MN and on the other hand ZGA, Achmea and SFK?

source 1	source 2	source 3	overlap
Achmea	JHN	PCR-MN	118140
SFK	JHN	PCR-MN	66671
ZGA	JHN	PCR-MN	213

Example 9.

In example 7 we have seen there are 38349 people found in both LifeLines and IADB. From IADB we have information about ATC codes. Therefore we can answer the question “How many LifeLines people are found in IADB and have diabetes?” Here we use ATC A10A and A10B as a proxy for diabetes.

Answer: we found 777 people .