



Lifebrain

D2.2. Enrichment of data by coupling to national registries

Project title:	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
Due date of deliverable:	31 th December, 2017
Submission date of deliverable:	22 nd December, 2017
Leader for this deliverable:	University of Oslo



This project has received funding from the *European Union's Horizon2020 research and innovation programme* under grant agreement No 732592.

Contributors	Name	Organisation	Role/ Title
Deliverable leader	Kristine B. Walhovd	UiO	Coordinator
Contributions	Barbara B. Friedman	UiO	Administrative coordinator
	Stine Kleppe Krogstad	UiO	Researcher
	David Bartrés-Faz	UB	PI
	Cristina Solé Padullés	UB	Researcher
	Brenda Penninx	VUmc	PI
	Laura Nawijn	VUmc	Researcher
	Lars Nyberg	UmU	PI
	Mikael Stiernstedt	UmU	Research coordinator
	Sara Pudas	UmU	
	William Baaré	RegionH	WP2 leader
	Final review and approval	Barbara B. Friedman	UiO

Document History				
Release	Date	Reason for Change	Status (Draft/In-review/Submitted)	Distribution
1.0	06.12.2017	First draft created	Sent for review to the Executive Board	E-mail, OneDrive
2.0.	22.12.2017	Final version	Final revision based on comments	Participant portal, OneDrive

Dissemination level		
PU	Public	X



Executive summary

While for some of the Lifebrain data, linkage to registry data will be of the nature of broader population comparisons, possibility of registry linkage for individualised data has been evaluated for some cohorts.

The Norwegian Lifebrain cohorts are already recruited through population registries and have already been linked by individual consent – e.g. the MRI cohorts from the Norwegian Mother and Child Study are linked, and Norwegian adult cohorts also have consented to linkage of their MRI, cognitive, mental health and genetic data with the Medical Birth registry as well as army conscription data.

The army conscription data of the Swedish Betula cohort are available for meta-analysis, but cannot be shared. Other registry data cannot be shared and not available for meta-analysis either at the moment, but might be available in the future.

The Dutch NESDA cohort can be linked to environmental and socio-economic registries, but not to birth registry.

As part of the Task2.3. Online data enrichment, self-reported information will be collected on birth and childhood.

Analysis of coupled data will be conducted in the related Work Package 4 tasks.

Table of contents

Executive summary	3
Table of contents.....	4
List of acronyms/ abbreviations	5
1. Introduction	6
1.1. Deliverable description	6
2. Possibilities of coupling of data to national registries	6
2.1. University of Oslo	6
2.1.1. Army conscription data	6
2.1.2. Variables from the Norwegian Birth Registry	7
2.2. RegionH.....	8
2.3. University of Barcelona	8
2.4. VU University Medical Centre	8
2.5. Umeå University.....	8
2.5.1. Army conscription data	9
2.5.2. Data from Statistics Sweden (Statistiska centralbyrån).....	9
2.5.3. Data from the National Board of Health and Welfare (Socialstyrelsen).....	10
3. Conclusion	14



List of acronyms/ abbreviations

Lifebrain	Healthy minds from 0-100 years: Optimising the use of European brain imaging cohorts
LCBC	Centre for Lifespan Changes for Brain and Cognition
LB	Lifebrain
M	Month
MoBa	Norwegian Mother and Child Study
NESDA	The Netherlands Study of Depression and Anxiety
REGIONH	Region Hovedstad
UB	University of Barcelona
UiO	University of Oslo
UmU	Umeå University
VUmc	VU University Medical Centre
WP	Work Package

1. Introduction

1.1. Deliverable description

D2.2 Data obtained from national registries; incl. standardization, integration, quality control and pooling with existing data across sites [12] Coupling of existing data with national health, conscription and population registries. This will be done for the Nordic countries (Norway, Sweden, Denmark), as well as the Netherlands and Spain.

Task 2.2.

Enrichment of data by coupling to national registries. Lead: UiO; Participants: UmU, UOXF, UB, REGIONH, VUmc (M6-M12) Coupling of existing data with national health, conscription and population registries. This will be done for the Nordic countries (Norway, Sweden, Denmark) as well as the Netherlands and Spain. For as many as possible of the participants in the samples from these countries, we aim to obtain information about birth weight and other neonatal health variables and general cognitive and physical function (e.g., by conscription data for males).

2. Possibilities of coupling of data to national registries

2.1. University of Oslo

Some of the Norwegian cohorts are already recruited through population registries and have already been linked by individual consent – e.g. the MRI cohorts from the Norwegian Mother and Child Study are linked, and Norwegian adult cohorts also have consented to linkage of their MRI, cognitive, mental health and genetic data with the Medical Birth registry as well as Army conscription data.

2.1.1. Army conscription data

- Data can be used for meta-analysis, but not shared
- Birth years of available participants approx. 1952-1996
- Number of observations (n) vary greatly between variables
- Height/weight are available for ~124 participants
- Physical measures (general physics, digestion, vision, hearing, arm, hand), measure of psychological health and cognitive test data are also available but with fewer observations (range from ~30- 96 observations)

2.1.2. Variables from the Norwegian Birth Registry

- See Table 1 for variables
- Data can be used for meta-analysis, but not shared
- Birth years of available participants approx. 1967-2007
- Birth variables available for the Norwegian adult cohorts
- MoBa is not included in the table because the participants (n ~ 521) are linked to all variables in the Norwegian birth registry

Table 1. Variables from the Norwegian Birth Registry

Variable	Variable name	English description	Available n
1	SVLEN	length of gestation in weeks (if SVLEN_UL_DG is given, if not SVLEN_SM_DG)	552
2	SVLEN_DG	length of gestation in days, composite	66
3	SVLEN_UL_DG	length of gestation in days, based on ultrasound examination	66
4	SVLEN_SM_DG	length of gestation, in days based on the woman's last menstrual period	549
5	KJONN	Sex	590
6	VEKT	Birth weight in grams	590
7	LENGDE	Birth length in centimeters	565
8	HODE	Head Circumference in centimeters	541
9	APGAR1	Apgare score after 1 minute	545
10	APGAR5	Apgare score after 5 minutes	543
11	APGAR10	Apgare score after 10 minutes	3

Summary: Linkage to Medical Birth registry and Army conscription data is possible, but will not be shared.

2.2. RegionH

It is possible to extract data on birth-registry and gestational age. However, given the relative small sample of typical-developing children, and the fact that data would need to be analysed on other research servers, a decision has been made not to pursue this. Some information on birthweight, and whether participants were born to early and whether there were birth complications along birth, exists. Moreover, we will try to retrieve this information for all participants through the planned Lifebrain data enrichment under Task 2.3.

Summary: Birth registry linkage is not possible.

2.3. University of Barcelona

Computerised clinical history is only available since 1996. For getting access to medical history on paper, a signed consent from each participant would be needed and papers would need to be collected from the different hospitals where people were born and ask for access to medical records in their files. It is not sure either that this data would be found in the medical records from 70 years back. Another issue is that many participants were not born in Barcelona city, but in small villages in Catalonia, Aragon and Andalusia, making it difficult to gather data. For the sample UB is following, participants had been asking about their birth weight, but only a handful of people could report on it.

Summary: Birth registry linkage is not possible.

2.4. VU University Medical Centre

Linkage to birth weight registry is not possible and no registry information is available on childhood either. However, environmental data are available on pollution (distance to water, air pollution, traffic noise) and socio-economic data on neighbourhood and safety are also available for coupling Lifebrain data with.

Summary: Birth registry linkage is not possible. Environmental and socio-economic data are available.

2.5. Umeå University

The following data are (partly) available in the Betula cohort for coupling.

2.5.1. Army conscription data

- Cognitive test data available for 432 men with at least one cognitive assessment in Betula – of which 80 have neuroimaging data (60 have two time-point neuroimaging data).
- Birth years of available participants approx. 1945-1955
- Data can be used for meta-analysis, but not shared
- Height/weight is NOT included in the data we have received from Krigsarkivet (would require additional data collection)

2.5.2. Data from Statistics Sweden (Statistiska centralbyrån)

- See variables in Table 2.
- Cannot be shared
- Will not be made available for meta-analyses at the moment – but may be available in the future
- Number of observations (n) vary greatly between files/variables, given numbers are approximations
- Betula participants are born between 1908 and 1985, so variables cover different portions of the lifespan for different individuals
- Some variables may have changed during the years, for instance the way in which income is calculated

Variable	Years	Additional info	Available n
Number of children			4445
Birth years of children			3824
Marital status	1968-2016		4258-4445
Family type	1990-2016	42 different categories (including info marital status, co-habitation, number of biological/adopted children etc.)	4440+
Income	1968-2015	Four types of income variables	2958-4445
Highest obtained educational degree	1990-2015	Two different categorizations available	~4400
Type of education obtained	1990-2015	Two variables, covering 93 and 199 categorical labels of types of education	~2747 - 4367
Occupation category	1990-2015	Mainly denotes entrepreneurs vs. "other employees"	~2747 - 4367
Detailed occupation category	2001-2015	Different categorizations ranging between 104-211 occupational categories	~2747 - 4367
Sickness benefit/compensation	1990-2015		~2747 – 4367
Unemployment benefit	1990-2015		~2747 – 4367
Early retirement pension	1990-2015		~2747 – 4367
Occupational pension	1990-2015		~2747 – 4367
Old age pension	1990-2015		~2747 – 4367

Table 2. Variables from the Swedish Statistical Bureau

2.5.3. Data from the National Board of Health and Welfare (Socialstyrelsen)

- See available registries in Table 3.
- See variables from the registries: Table 4-Table 8.
- Cannot be shared
- Will not be made available for meta-analyses at the moment – but may be available in the future

Registry	Years	Number of variables	Number of observations
Cancer registry	1960-2015	28	1585
Cause of death registry	1990-2016	64	1760
Drug prescription registry	2004-2016	23	934796
Patient registry, outpatient care	2001-2015	16	70562
Patient registry, inpatient care	1964-2015	20	33436

Table 3. Registry data available from the National Board of Health and Welfare

Variables from the cancer registry (Table 4):

Variable	Variable name	English description
1	KON	Sex
2	SJUKHUS	Hospital
3	KLINIK	Clinic
4	DIADAT	Date of diagnosis
5	ICDO3	Tumour localization according to International Classification of Diseases for Ontology (v.3)
6	ICDO10	Tumour localization according to International Classification of Diseases for Ontology - version 2-10
7	ICD9	Tumour localization according to International Classification of Diseases (1987-1996)
8	ICD7	Tumour localization according to International Classification of Diseases (1958-1963, alt. 1964-1968)
9	SIDA	Side (of body)
10	T	Size of primary tumour (TNM-classification)
11	N	Regional lymph node involvement (TNM-classification)
12	M	Distant metastasis (TNM-classification)
13	FIGO	Tumour extent at time of diagnosis, gynaecological tumours
14	SNOMED3	Morphological diagnosis of tumour according to ICD-O/3
15	SNOMEDO10	Morphological diagnosis of tumour according to ICD-O/2
16	PAD	Histopathological diagnosis of tumour according to C24.1
17	ben	Indicator for malignant or benign tumour according to morphology
18	OBD1	Tumour identified at autopsy
19	TNMGRUND	Examination that TNM is based on
20	ulorsak	Disease or injury that initiated chain of events leading to death
21	dbgrund1	Source for cause of death. If clinical or forensic autopsy.
22	Diadatn	Date of confirmation of diagnosis
23	Foddat	Date of birth, year month
24	Ar	Year of diagnosis
25	Alder	Age at diagnosis
26	Tnr	Tumour number, chronologically, if several
27	Tnrml	Tumour number, chronologically, only malignant tumours
28	LopNr	Identification number

Table 4. Variables from the cancer registry

Variables from the cause of death registry (Table 5):

Variable	Variable name	English description
1	DODSDATN	Date of death
2	FODDAT	Birthd date, year month
3	DODSDAT	Date of death
4	KON	Sex
5	Alder	Age at death
6	LopNr	Identification number
7	ICD	Version of ICD (International classification of diagnoses)
8	LKF	County, municipality
9	ULORSAK	Underlying cause of death
10	KAP19	Main injury or poisoning
11	MORSAK1	1st multiple cause of death
12	MORSAK2	2nd multiple cause of death
13	MORSAK3	2rd multiple cause of death
14	MORSAK4	4th multiple cause of
15-58	MORSAK5	5th-48th multiple cause of death
59	DBGRUND1	If clinical or forensic autopsy was performed
60	ALKOHOL	If alcohol-related diagnosis mentioned on the certificate, either underlying or contributing cause of death
61	DIABETES	If diabetes was either underlying or contributing cause or death
62	FCIV	Preceding marital status
63	KAP17	When underlying cause of death is external, code from ICD 7/8/9
64	UBASIC	Basic Tabulation List, ICD 9

Table 5. Variables from the cause of death registry

Variables from the drug prescription registry (Table 6):

Variable	Variable name	English description
1	AR	Year of visit
2	LKF	County, municipality
3	FODDAT	Date of birth, year month
4	KON	Sex
5	ALDER	Age at visit
6	SJUKHUS	Hospital
7	MVO	Type of medical specialty
8	INDATUMA	Date of contact with outpatient care center
9	INDATUM	Date of contact with outpatient care center, numeric
10	KTYP	Type of contact (e.g. Home/clinic, group/individual, phone etc.)
11	HDIA	Main diagnosis according to classification by Sosialstyrelsen
12	DIAGNOS	Diagnoses according to ICD10, max 30
13	OP	Type of action/intervention
14	CIVIL	Marital status of patient
15	LopNr	Identification number
16	LT_KLIN	Clinic for visit

Table 6. Variables from the drug prescription registry

Variables from the patient registry for outpatient care (Table 7):

Variable	Variable name	English description
1	AR	Year of visit
2	LKF	County, municipality
3	FODDAT	Date of birth, year month
4	KON	Sex
5	ALDER	Age at visit
6	SJUKHUS	Hospital
7	MVO	Type of medical specialty
8	INDATUMA	Date of contact with outpatient care center
9	INDATUM	Date of contact with outpatient care center, numeric
10	KTYP	Type of contact (e.g. Home/clinic, group/individual, phone etc.)
11	HDIA	Main diagnosis according to classification by Sosialstyrelsen
12	DIAGNOS	Diagnoses according to ICD10, max 30
13	OP	Type of action/intervention
14	CIVIL	Marital status of patient
15	LopNr	Identification number
16	LT_KLIN	Clinic for visit

Table 7. Variables from the patient registry for outpatient care

Variables from the patient registry for inpatient care (Table 8):

Variable	Variable name	English description
1	AR	Year of discharge
2	LKF	County, municipality
3	FODDAT	Date of birth, year month
4	KON	Sex
5	ALDER	Age at discharge
6	SJUKHUS	Hospital
7	MVO	Type of medical specialty
8	INDATUMA	Date of admission
9	UTDATUMA	Date of discharge
10	INDATUM	Date or admission, numeric
11	UTDATUM	Date or discharge, numeric
12	VTID	Number of days between admission and discharge
13	INSATT	Type of admission
14	UTSATT	Type of discharge
15	HDIA	Main diagnosis
16	DIAGNOS	Diagnoses according to ICD10, max 30
17	OP	Type of action/intervention according to classification by Socialstyrelsen, max 30
18	CIVIL	Marital status of patient
19	LopNr	Identification number
20	LT_KLIN	Clinic for visit

Table 8. Variables from the patient registry for inpatient care

Summary: Army conscription data are available for meta-analysis, but cannot be shared. Other registry data will not be made available for meta-analyses at the moment – but may be available in the future.

3. Conclusion

Possibility of coupling of existing data with national health, conscription and population registries has been checked for the cohorts in Norway, Sweden, the Netherlands, Denmark and Spain. Where possible - Army conscription data and Medical Birth Registry in Norway, Army conscription data in Sweden, environmental and socio-economic data in the Netherlands -, Lifebrain data will be coupled with national health, conscription and population registries and meta-analysis will be done in WP4 Demonstration.