

GRELL & ENCR

2025 Scientific Meeting Porto, Portugal

27-30 MAY



Privacy-preserving self-service linkage of cancer registry and socioeconomic data in a safe processing environment

Siri Larønningen

Cancer Registry of Norway, Norwegian Institute of Public Health



Background

Problems:

- Time consuming applications
- Lacking metadata
- Lacking possibility to explore data before the application (“cohort-explorer”)
- Physical sharing of pseudonymized data
- Limited control over the data after it has been shared

→ It isn't always necessary to see the actual individual data

One possible solution: self-service in a secure environment

microdata.no – data access without application



microdata.no provides instant, online access to large amounts of detailed and mergeable microdata without any form of application.

The service is open to employees and students at universities and colleges, approved research institutions, ministries and directorates.

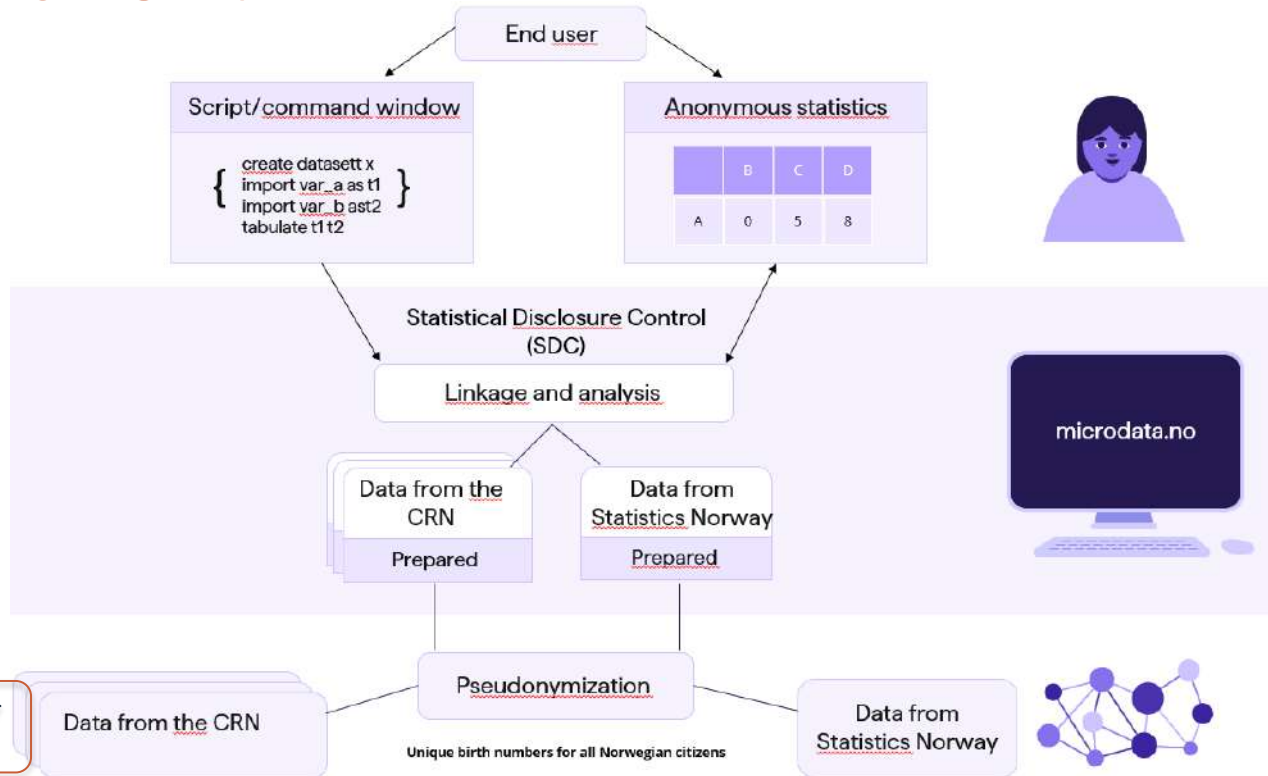
- No applications
- Instant access
- Time series from 1964
- Export function. Create a dataset and apply for it.
- Self-service. The institutions register their users themselves.

A solution supporting the FAIR-principles!

Microdata.no

- Has been in use with data from Statistics Norway since spring 2018
 - Currently more than 700 users from 75 institutions
 - 350 publications have used statistics made in microdata.no (level 2 publications, PhDs, reports, reviews, master thesis etc.)
 - 665 variables from Statistics Norway is currently available
-
- AND: From Autumn 2025, variables from the Cancer Registry of Norway and from the Norwegian Patient Registry will also be available

How does it work?



Metadata

SEER stage

Add to my variables

Metadata X

Hide more search options

Found: 40

Name

Morphology ICD03

Tumor type, the histopathological classification of tumor's digit. Differentiation of tumor is the sixth digit.

Pathology stage

Stage indicated on pathology report. Exception: Stage indicated on clinical report for gynecological cancer.

Person status

Information of whether the patient is alive and living in Norway.

pM

Pathological M-stage.

pN

Pathological N-stage.

pT

Pathological T-stage.

Radiotherapy

Information about radiotherapy.

SEER stage

Three-part summary stage which is consistent with stage as given in U.S. National Cancer Institute's Surveillance, Epidemiology and End Results (SEER) database.

Metadata X

Hide more search options

Found: 40

Name

Age at diagnosis

Age of the patient at time of diagnosis.

Basis of diagnosis

The most reliable method used to confirm diagnosis.

Birthnumber

The patient's national identification number.

Cancer case in Cancer in Norway

Indicates whether the cancer case is included in the Cancer in Norway database.

Categorization of underlying cause of death

Describes whether the patient is dead because of cancer.

Cause code for underlying cause of death

If underlying cause of death is cancer, the cause of death is given according to these classifications: 1951-1957: ICD-6 1958-1968: ICD-7 1969-1985: ICD-8 1986-1999: ICD-9

Certainty of diagnosis

How reliable the diagnosis of the cancer case is. The reliability is assessed based on certainty of tumor malignancy and certainty of primary site of tumor.

Overview

Cancer sites

Clinical registries

Values

References

Detail

Quality

Value

0

Value description

Benign/In situ

1

Localised

2

Regional

3

Distant metastasis

9

Unknown

Topography ICD03

Overview

Cancer sites

Clinical registries

Values

References

Detail

Quality

Se informasjon om ICD-O-3 på IARC/WHO sine nettsider: http://www.iacr.com/fr/index.php?option=com_content&view=category&layout=blog&id=100&Itemid=1

Value

000

Value description

External lip upper

Valid from

01-01-1953

Valid to

Now

Cancer sites

C00

001

External lip lower

01-01-1953

Now

C00

002

External lip NOS

01-01-1953

Now

C00

003

Upper lip, mucosa

01-01-1953

Now

C00

004

Lower lip, mucosa

01-01-1953

Now

C00

005

Mucosa lip, NOS

01-01-1953

Now

C00

006

Commissure lip

01-01-1953

Now

C00

008

Overlapping lesion of lip

01-01-1953

Now

C00

009

Lip nos

01-01-1953

Now

C00

Already in my variables

```

1 //Connect to KRG-databank (DRAFT = available for datatesters and dataadmins
  only)
2 require no.fhi.kreftregisteret:DRAFT as kreg
3 //Connect to Statistics Norway-databank
4 require no.ssb.fdb:34 as fdb
5
6 //Cancer incident dataset
7 create-dataset cancer_incident
8
9 //Importing date of diagnosis
10 // import kreg/S_DIAGNOSEDATO as diagdato
11 import kreg/S_DIAGNOSEDATO as diagdato
12
13 //Using the year()-function to extract year from epoch days in variable
14 // generate diagnoseyear = year(diagdato)
15 generate diagnoseyear = year(diagdato)
16
17 //Frequency table of cancer incidents pr year
18 tabulate diagnoseyear
19
20 //Keeping only incidents from 2022 for now
21 keep if inrange(diagnoseyear, 2022, 2022)
22
23 //Import S_SEERSTADIUM (Summarisk tredeling av stadium i samsvar med U.S.
  National Cancer Institute's SEER Program)
24 import kreg/S_SEERSTADIUM as seerstadium
25
26 //Import av "krefttilfellets fødselsnummer (pseudonymt)"
27 import kreg/P_FODSELSNR as kreg_fnr
28
29 //Person dataset with data from SSB
30 create-dataset person
31 // *** RESIDENTS ONLY ***
32 //Importing the STATUSCODE variable that lets us
33 //keep only the population that lives in the country on
34 //a given date
35 import fdb/BEFOLKNING_STATUSCODE 2022-01-01 as regstat22
36 keep if regstat22 == '1'
37
38 // *** EDUCATION ***
39 //Import highest education by 2022
40 import fdb/NUDB_BU 2022-01-01 as highest_education_22
41

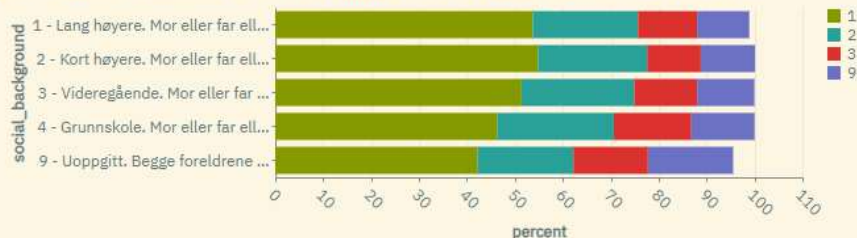
```

Test data – not for publication

social_background

	seerstadium				
	1	2	3	9	Total
1 - Lang høyere. Mor eller far eller begge har utdanning på nivå 7 el. 8	53.74	21.96	12.31	10.9	100
2 - Kort høyere. Mor eller far eller begge har utdanning på nivå 6	54.85	22.75	11.2	11.32	100
3 - Videregående. Mor eller far eller begge har utdanning på nivå 3, 4 eller 5	51.39	23.53	13.13	11.94	100
4 - Grunnskole. Mor eller far eller begge har utdanning på nivå 0, 1 eller 2	46.31	24.35	16.18	13.19	100
9 - Uoppgitt. Begge foreldrene har uoppgitt utdanning.	42.22	20	15.56	17.78	100
Total	49.83	23.68	14.13	12.37	100

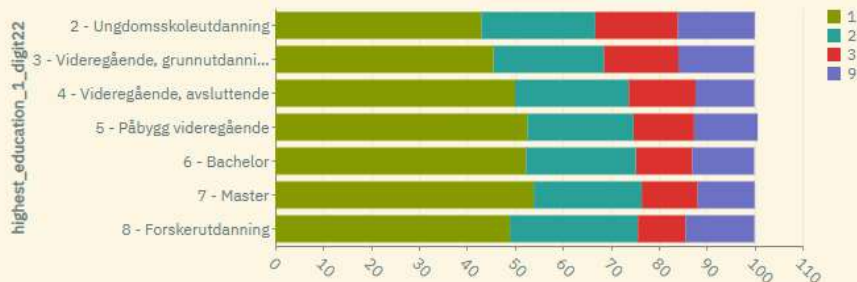
cancer_incident>barchart (percent) seerstadium, over(social_background) stack horizontal



cancer_incident>drop if inlist(highest_education_1_digit22, "0", "1")

176 enheter ble fjernet fra datasettet.

cancer_incident>barchart (percent) seerstadium, over(highest_education_1_digit22) stack horizontal



Security in microdata.no

The Five Safes framework:

Safe data: data is treated to protect any confidentiality concerns

Safe projects: research projects are approved by data owners for the public good.

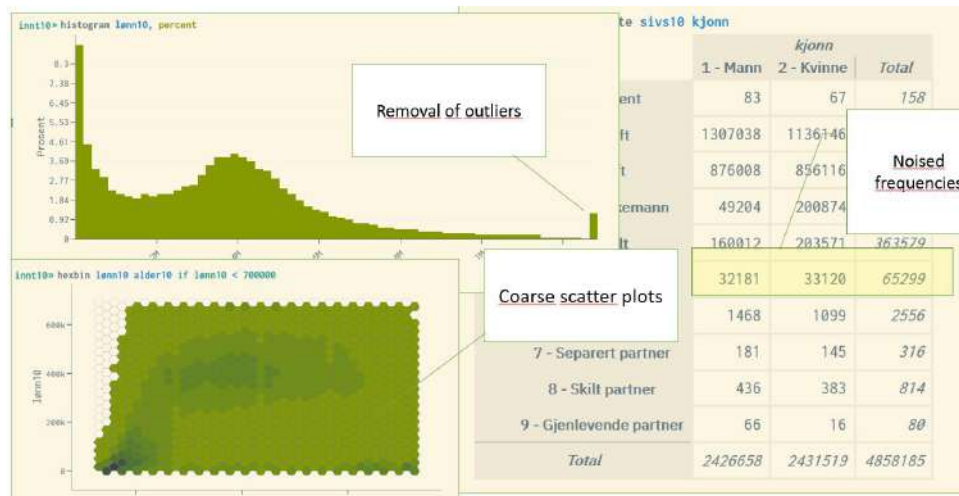
Safe people: researchers/users are trained and authorised to use data safely.

Safe settings: the environment prevents unauthorised use.

Safe outputs: screened and approved outputs that are non-disclosive.



Statistical Disclosure Control (SDC)



Misuse detection



USERS CAN BE REPORTED OR
BLOCKED, BASED ON
BEHAVIOR



SYSTEM ADMINISTRATORS
CAN CHECK THE
REPORTED/BLOCKED SCRIPT
AND DECIDE ON ACTIONS



ACTIONS CAN BE WHITE-
LISTED OR BLACK-LISTED



UTILITY VS. PRIVACY
DISCUSSIONS



Pros and cons of microdata.no

PROs

- Instant accessibility if authorised user
- Linkage of health data and socioeconomic data
- Cohort-explorer
- Scripts can be shared and re-used (for instance with an application or individual data)
- Privacy-preserving

CONs

- Need to know how to write scripts and analyze data
- Not accessible for everyone (for instance drug companies, private companies etc)
- Limitations due to the SDC-layer
- Not suitable for complex analysis where direct access to data is crucial

**Big thanks to the super-cool purple
team from Statistics Norway, the
Norwegian Agency for Shared Services
in Education and Research, and the
Cancer Registry of Norway**

Statistics Norway:

Rune Gløersen, Vidar N. Klungre

Norwegian Agency for Shared Services in Education and Research:

Ørnulf Rises, Sigbjørn Revheim, Eirik Alvær, Eirik B. Stavestrand

Cancer Registry of Norway:

Jan F. Nygård, Sigrid Leithe, Narasimha Raghavan, Gintaras Pikelis,
Siri Larønningen

