European Network of Cancer Registries

Workshop 4

Best Practices for Implementing Quality Check Software and Reporting Tools: Lessons Learnt from NORDCAN, and future direction

Siri Larønningen, Cancer Registry of Norway

Granada, Spain, November 13th, 2023

Cancer Registry of Norway

International Agency for Research on Cancer
World Health Organization

NORDCAN
Association of the Nordic Cancer Registries

NORDIC CANCER UNION

European Commission

Pick a data visualization

Incidence/Mortality

- Trends
- Age-specific
- Cohorts
- Maps
- Bar charts
- Tables
- EAPC
- Population pyramids

Prevalence

- Prevalence - Trends
- Prevalence - Tables

Survival

- Survival - Trends
- Survival Age Specific - Trends (New)
- Survival - Bar charts
- Survival Age Specific - Bars (New)
- Survival - Tables
- Survival improvements

- Incidence/Mortality
- Prevalence
- Survival
- Predictions
- Quality tables (New)

Predictions

- Predictions - Trends
- Predictions - Cohorts
- Predictions - Age specific (New)
- Predictions - Bars
- Predictions - Tables

Quality tables

- Quality tables (Incidence) (New)
- Quality tables exclusion (Incidence) (New)

European Network of Cancer Registries

Cancer Registry of Norway

International Agency for Research on Cancer
World Health Organization

NORDCAN
Association of the Nordic Cancer Registries

NCU NORDIC CANCER UNION

European Commission

Display | Options

Home / Data visualization / Trends

Graphic

Display by
Population

Statistics
ASR (Nordic)

Measures
Incidence | Mortality | Inc./Mortality

Sexes
Males | Females | By sex | **Both sexes**

Cancer sites (1)
Grouped

All sites

Populations (7)
Grouped

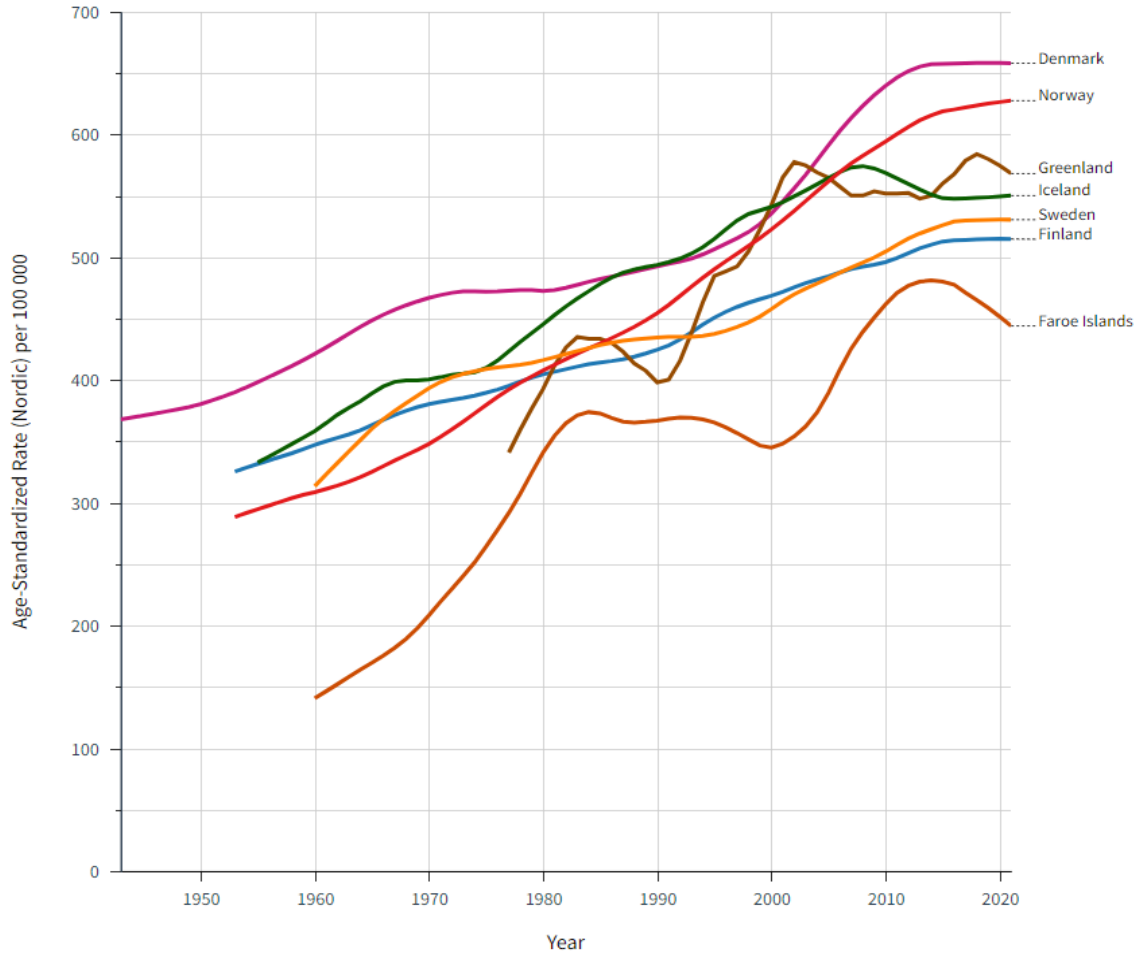
Denmark × Finland × Iceland × Norway × Sweden × Faroe Islands × Greenland ×

Age groups
0 — 85+

Years
1943 — 2021

Re-initialize all values

Age-Standardized Rate (Nordic) per 100 000 , Incidence, Both sexes
All sites
Denmark - Faroe Islands - Finland - Greenland - Iceland - Norway - Sweden



Lines are smoothed by the LOESS regression algorithm (bandwidth: 0.2)

9.3

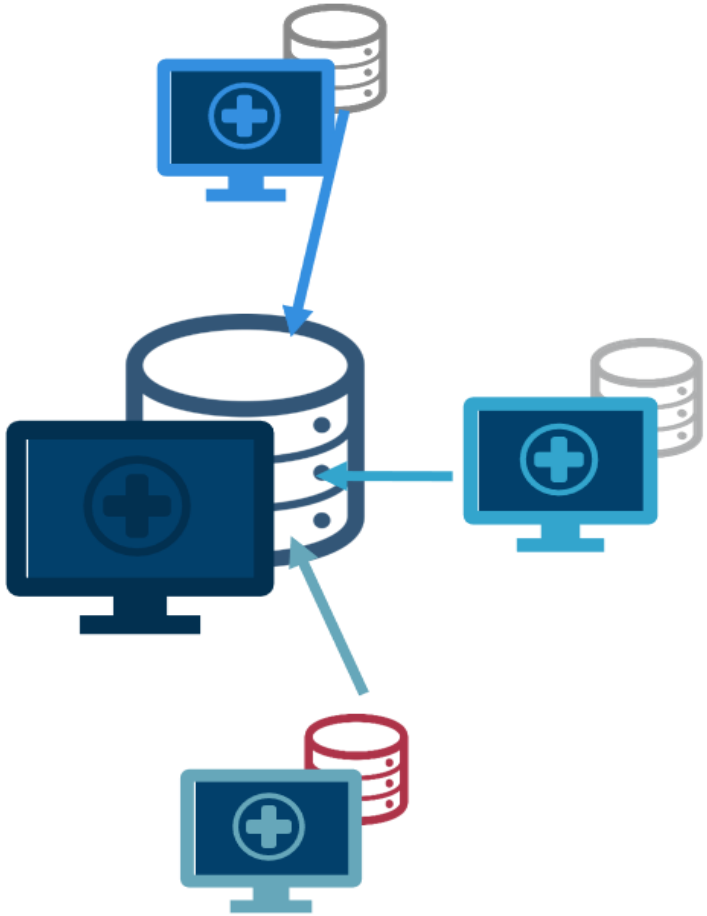
European Network of Cancer Registries
Cancer Registry of Norway
International Agency for Research on Cancer / World Health Organization
NORDCAN Association of the Nordic Cancer Registries
NCU Nordic Cancer Union
European Commission

GDPR and legal issues

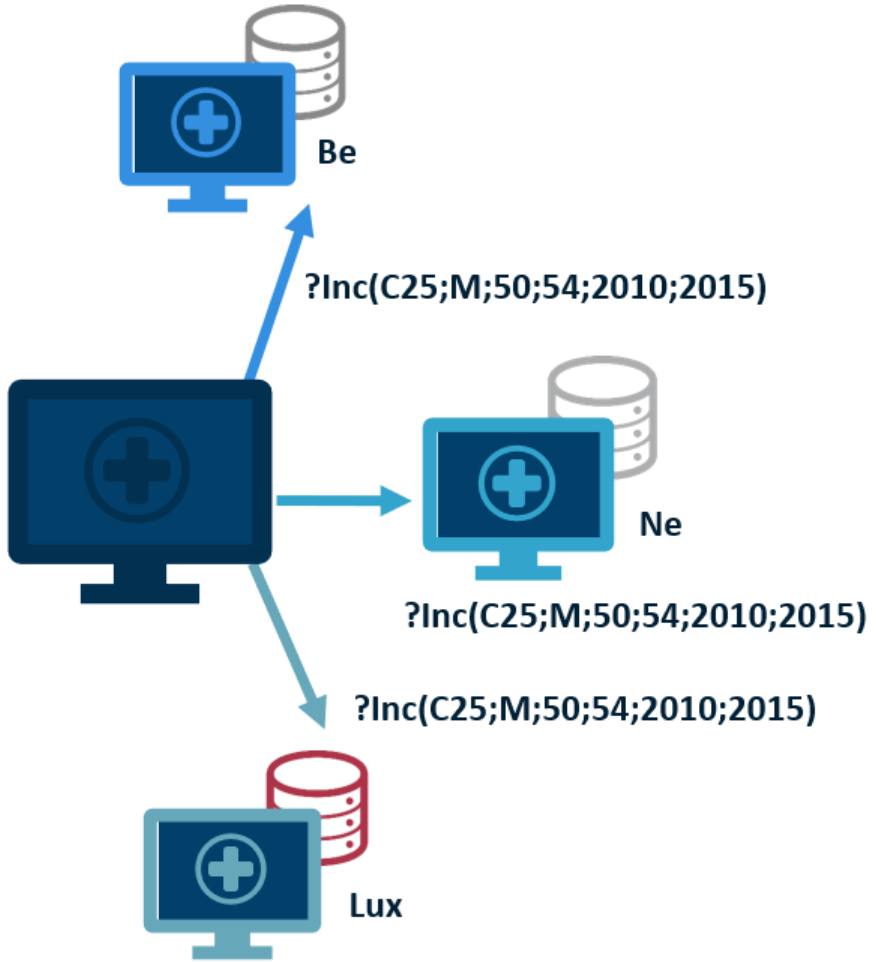
- In effect since 2018
- Strengthened standards for sharing personal data
- Country-specific legislation/regulations
  - Data minimization
  - Broad purposes

- We could now only share anonymous data with IARC
- We had to rethink our data flow!

Centralized approach



Federated approach



The basics

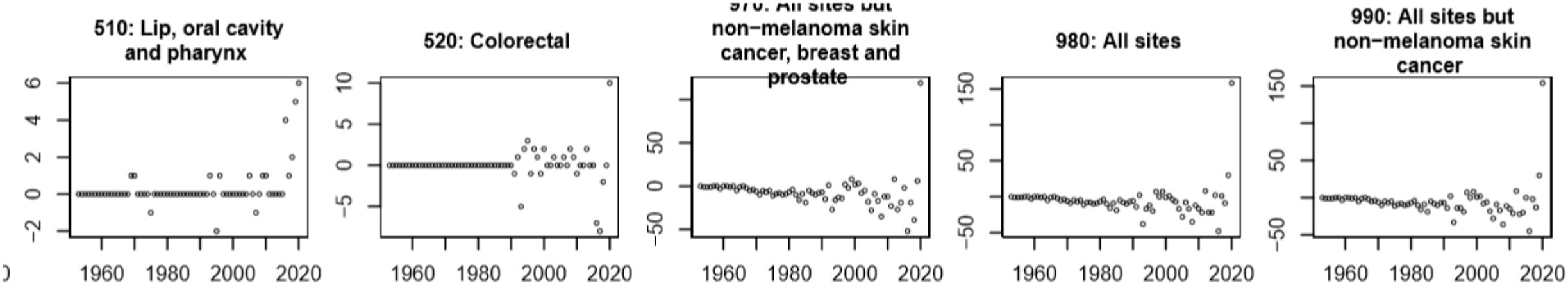
- Common data model
- Harmonized standards (and practices)
- Nordcan.R: R-scripts + Stata + IARCcrgTools
- Definition of the output
- Means for quality control

```
####################################
## NORDCAN data processing starts ...
####################################

## The processing includes 5 sections:
## (1) Manually specify the paths of 'IARCcrgTools', 'STATA',
##     raw data (cancer record, etc.), and set up NORDCAN global settings
## (2) Import raw data into R, and pre-process the cancer record (with IARCcrgTools)
##     & cancer death count data.
## (3) Generate the NORDCAN statistics tables. This section is time consuming.
## (4) Compare the new-calculated statistics tables with an older version.
## (5) Save result for archive and sending.

## To run survival analysis, the disk where STATA is installed should has 2GB
## free space at the minimum!

#########################################################
## (1) Manually specify the paths & NORDCAN global settings

path_IARC  <- "C:\\Program Files (x86)\\IARCcrgTools\\IARCcrgTools.EXE"
path_STATA <- "S:\\Prog64\\Stata\\Stata18MP\\StataMP-64.exe"

## paths of raw dataset
file_incidence  <- "P:\\Dataflyt\\nordcan\\data\\2021\\incidence_2021.csv"
file_lifetable  <- "P:\\Dataflyt\\nordcan\\data\\2021\\national_population_life_table_2021.csv"
file_population <- "P:\\Dataflyt\\nordcan\\data\\2021\\population_2021.csv"
file_mortality  <- "P:\\Dataflyt\\nordcan\\data\\2021\\mortality_2021.csv"

## path of population projection data
## (if there is no such file, just leave it as it is).
file_pop_proj   <- "P:\\Dataflyt\\nordcan\\data\\2021\\population_projection_2021.csv"

## directory for saving the output of NORDCAN processing.
dir_result <- "P:\\Dataflyt\\nordcan\\user\\ansk\\nordcan_9.3_1.3\\results"

## directory for holding the archived (.zip) result.
dir_archive <- "P:\\Dataflyt\\nordcan\\user\\ansk\\nordcan_9.3_1.3\archive"

## path of previous archived statistics result (.zip) to compared with.
stats_archived <- "P:\\Dataflyt\\nordcan\\archive\\nordcan_9.2_statistics_tables.zip"

## Set up global settings.
## Remember to modify the 'participant_name' and 'last_year_..'
nordcancore::set_global_nordcan_settings(
  work_dir = dir_result,
  participant_name = "Norway", # need to be modified
  first_year_incidence = 1953L,
  first_year_mortality = 1953L,
  first_year_region    = 1953L,
  last_year_incidence  = 2021L,
  last_year_mortality  = 2021L,
  last_year_survival   = 2021L
)
```

Steps of the script

- Install packages, set basic settings
- Read files, set country-specific and on-prem settings
- Quality assure and enrich
  (through IARCcrgTools and other functions)
- Run analysis
- Compare to archived file
- Export zip-file with output

Example of quality control tools



Pros and cons

Advantages

- Sharing only anonymous data with IARC
- GDPR compliant: no individual data leaves the CRs
- User can immediately check results and quality assure and, if necessary, make new, corrected input data
- Data managers who know their data well
- Building on core CR variables with long history of use and standardisation

Disadvantages

- Not a «real» federated setup
- No pooling of data
- Mainly suited for less complex datasets
  - Establishing a common ground for many variables is difficult
- Strict rule of anonymity in output

Building the foundation

Nordcan.R is a pragmatic first step

Foundation for further development

- Federated analysis
- Federated learning
- Synthetic data

The Nordic countries have:

- population of above 27 million people (2021)
- 70-year history of quite harmonized cancer registration.

Building the future

BlueBerry
blueprint for the EURACAN registry

European collaboration in FLORENCE project Using AI to improve treatment of patients with colorectal cancer

IDEA4RC
Intelligent ecosystem to improve the governance, the sharing, and the re-use of health data for rare cancers

NORDICRWE
REAL WORLD EVIDENCE

frontiers | Frontiers in Oncology

TYPE Technology and Code
PUBLISHED 08 August 2023
DOI 10.3389/fonc.2023.1098342

OPEN ACCESS

EDITED BY
Liesbet Van Eycken,
Belgian Cancer Registry, Belgium

REVIEWED BY
Stephanie G. C. Kroeze,
Aarau Cantonal Hospital, Switzerland
Roberta De Angelis,
National Institute of Health (ISS), Italy
Emanuele Crocetti,
Independent researcher, Florence, Italy
Maximilian Knoll,
German Cancer Research Center (DKFZ), Germany

*CORRESPONDENCE
Siri Larønningen
sla@kreftregisteret.no

RECEIVED 14 November 2022
ACCEPTED 04 July 2023
PUBLISHED 08 August 2023

CITATION
Larønningen S, Skog A, Engholm G,
Ferlay J, Johannesen TB, Kristiansen MF

Nordcan.R: a new tool for federated analysis and quality assurance of cancer registry data

Siri Larønningen[1*], Anna Skog[1], Gerda Engholm[2], Jacques Ferlay[3], Tom Børge Johannesen[1], Marnar Fríðheim Kristiansen[4], Daan Knoors[5], Simon Mathis Kønig[2], Elinborg J. Olafsdottir[6], Sasha Pejicic[7], David Pettersson[7], Charlotte Wessel Skovlund[2], Hans H. Storm[2], Huidong Tian[1], Bjarte Aagnes[1] and Joonas Miettinen[8]

[1]Department of Registration, Cancer Registry of Norway, Oslo, Norway, [2]Cancer Surveillance and Pharmacoepidemiology, Danish Cancer Society Research Center, Copenhagen, Denmark, [3]Cancer Surveillance Branch, International Agency for Research on Cancer, Lyon, France, [4]University of the Faroe Islands, Faculty of Health Sciences, Tórshavn, Faroe Islands, [5]Research and Development, Netherlands Comprehensive Cancer Organisation, Eindhoven, Netherlands, [6]The Icelandic Cancer Registry, Reykjavik, Iceland, [7]The National Board of Health and Welfare, Stockholm, Sweden, [8]Finnish Cancer Registry, Helsinki, Finland

Aim of the article: We present our new GDPR-compliant federated analysis programme (nordcan.R), how it is used to compute statistics for the Nordic cancer statistics web platform NORDCAN, and demonstrate that it works also with non-Nordic data.

Materials and methods: We chose R and Stata programming languages for writing nordcan.R. Additionally, the internationally used CRG Tools programme by International Agency for Research on Cancer (IARC/WHO) was employed. A formal assessment of (GDPR-compliant) anonymity of all nordcan.R outputs was performed. In order to demonstrate that nordcan.R also works with non-Nordic data, we used data from the Netherlands Cancer Registry.

Results: nordcan.R, publicly available on Github, takes as input cancer and general population data and produces tables of statistics. Each NORDCAN participant runs nordcan.R locally and delivers its results to IARC for publication. According to our anonymity assessment the data can be shared with international organizations, including IARC. nordcan.R incidence results on Norwegian and Dutch data are highly similar to those produced by two other independent methods.

Conclusion: nordcan.R produces accurate cancer statistics where all personal and sensitive data are kept within each cancer registry. In the age of strict data

https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2023.1098342/full

www.encr.eu