

NORDCAN risk assessment

Version 1.0
NORDCAN, December 2020

Contents

1	RISK ASSESSMENT SCORING	4
1.1	SCORECARD FOR RISK ASSESSMENT	4
1.2	ACCEPTABLE AND NON-ACCEPTABLE RISK	5
1.3	ASSESSMENT OF RISK	5
2	DEFINITIONS.....	6
3	CONTENTS OF THE NORDCAN DATABASE AND WEB TOOL	7
3.1	INCIDENCE, PREVALENCE, MORTALITY AND POPULATION	7
3.2	SURVIVAL.....	8
3.3	TIMELINESS AND COMPLETENESS OF DATA	8
3.3.1	TIMELINESS.....	8
3.3.2	COMPLETENESS	8
4	NORDCAN RISK ASSESSMENT	9
4.1	AGGREGATE COUNTS BASED ON NUMBERS IN NORDCAN FROM 1943 TO 2016.....	9
4.1.1	SMALLEST NUMBER OF PERSONS AT RISK.....	9
4.1.2	SMALLEST NUMBER OF CASES – SMALLER COUNTRIES.....	11
4.1.3	SMALLEST NUMBER OF CASES – LARGER COUNTRIES.....	12
4.2	RISKS OF DISCLOSURE OF DATA CONCERNING HEALTH THROUGH NORDCAN AGGREGATE COUNTS AND STATISTICS	13
4.3	DISADVANTAGES OF NOT PUBLISHING NORDCAN AGGREGATE COUNTS AND STATISTICS	15
4.4	CONCLUSION OF THE RISK ASSESSMENT.....	16

5	<u>RISK-REDUCING MEASURES.....</u>	<u>17</u>
5.1	RISK-REDUCING MEASURES ALREADY IMPLEMENTED IN NORDCAN	17
5.1.1	RISK REDUCING MEASURES FOR AGGREGATE COUNTS AND STATISTICS IN NORDCAN.....	17
5.2	ACTION PLAN FOR RISKS IDENTIFIED THROUGH THE RISK ASSESSMENT.....	18
5.3	SHARING AGGREGATE COUNTS AND STATISTICS WITH IARC	20
6	<u>LITERATURE</u>	<u>21</u>

1 Risk assessment scoring

1.1 Scorecard for risk assessment

The risk assessment scores different scenarios according to this scale:

Likelihood of the scenario happening

- Low (1)
- Medium (2)
- High (3)
- Very high (4)

Consequence or impact if scenario happens

- Low (1)
- Medium (2)
- High (3)
- Very high (4)

Combined, this gives the following matrix:

Likelihood	Very high	Medium risk (4)	High risk (8)	Critical risk (12)	Critical risk (16)
	High	Medium risk (3)	High risk (6)	Very high risk (9)	Critical risk (12)
	Medium	Low risk (2)	Medium risk (4)	High risk (6)	High risk (8)
	Low	Low risk (1)	Low risk (2)	Medium risk (3)	Medium risk (4)
		Low	Medium	High	Very high
Consequence / Impact					

1.2 Acceptable and non-acceptable risk

Based on the scorecard, the combination of Likelihood and Consequence/Impact gives us a risk score which will be used to evaluate whether the risk is acceptable. Non-acceptable risks need to be handled through risk-reducing strategies (see document “NORDCAN – protecting confidentiality in tables and graphs of cancer statistics in the Nordic countries”) and then re-assessed.

Risk	Acceptability
Low risk (score 1, score 2)	Acceptable
Medium risk (score 3, score 4)	Acceptable
High risk (score 6, score 8)	Attention, to be assessed
Very high risk (score 9)	Non-acceptable, risk-reducing strategies need to be implemented
Critical risk (score 12, score 16)	Non-acceptable, risk-reducing strategies need to be implemented

1.3 Assessment of risk

The tables used for risk assessment in NORDCAN consist of the following items:

- **#:** Unique number tied to a single possible threat/scenario
- **Scenario:** A short description of the possible threat/scenario that is to be assessed
- **Possible consequence:** A short description of the possible consequence of the scenario
- **C:** Assessed consequence (1-4) if the scenario happens
- **L:** Assessed likelihood (1-4) of the scenario happening
- **Score:** The product of consequence and likelihood
- **Justification:** Written justification of the assessed consequence and likelihood
- **Acceptability:** Whether the risk of the scenario is acceptable or non-acceptable
- **Action #:** Unique number for each measure to address non-acceptable risk

2 Definitions

Term	Definition
Counts / aggregate counts	The number of cancer cases or cancer deaths within a group. The result of a calculation/computation/aggregation of data records to aggregate numbers. When calculating rates, the counts of cancer cases/cancer deaths are the numerators, and the counts of population are the denominators
Cancer site	Used as a general term for cancer type, for instance breast cancer, prostate cancer, colorectal cancer, melanoma
Entity	Cancer site or group of cancer sites
Data subject	An identifiable natural person as defined in the GDPR, article 4, nr. 1 (1)
Data concerning health	Personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status. Defined in GDPR article 4, nr. 15 (1)
Persons at risk	The number of data subjects that can be subject to the given risk, for instance that all males in Norway are persons at risk for at prostate cancer and all women in Norway are persons at risk for a cervical cancer. In cancer statistics frequently used for the number of data subjects within a specific sex, age group and geographical localization. "Persons at risk" is the denominator in the calculation of rates.

3 Contents of the NORDCAN database and web tool

3.1 Incidence, prevalence, mortality and population

The NORDCAN database and web tool contains the following information for incidence, prevalence and mortality.

Variable	Content
Age at time of diagnosis	5 year age groups from 0-4 to 85+ (old version) or 90+ (new version). See overview here: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Specification-Age-groups
Current sex	The current sex of the patient, either “male” or “female”
Year of diagnosis	The year the cancer case was confirmed, i.e. the diagnostic year of each cancer case
Entity	Cancer site or grouping of cancer sites. The entities used in NORDCAN can be seen here: https://metadata.kreftregisteret.no/variables/detail/814?tabIndex=4 and here: https://nordcan.iarc.fr/en/database
Region/country	Region and/or country where the patient was resident at the time of diagnosis. See overview here: https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Specification-Regions

Only aggregate counts are sent to IARC for incidence, prevalence and mortality, not individual records. The typical cancer case file looks like this:

Entity	Agegroup	Year	Sex	Region	Cancer_record_count
10	12	2016	1	51	40
10	13	2016	1	51	25
...
980	18	2016	1	54	700
980	18	2016	2	54	800

The mortality files have the same structure as the cancer case file, except that the cancer record count is replaced by a cancer death count, meaning that we provide the number of deaths with a certain entity as underlying (main) cause of death within a specific year, age group, region and sex.

The population files give the population count at mid-year¹ and at the end of year for each region, age group, sex and year. These files are downloaded from the national statistics bureaus and used for the calculation of rates. IARC is also provided with population projections from the national statistics bureaus for future years to be able to calculate predictions.

3.2 Survival

For survival, no counts are transferred to IARC – only calculated estimates for survival based on well-defined statistical methods. More information on survival methods used in NORDCAN can be found at <https://github.com/CancerRegistryOfNorway/NORDCAN/wiki/Module-Survival>.

3.3 Timeliness and completeness of data

3.3.1 Timeliness

The timeliness of data in NORDCAN is a product of the timeliness of the cancer registries providing data for NORDCAN. Complete data is usually not published from the cancer registries until, at earliest, 9-10 months after the end of the diagnostic year. Data in NORDCAN cannot currently be used to look at the most recent (current year) statistics. Cause of death is typically received with a delay and therefore complete mortality data are available only with a delay of one to two full calendar years.

3.3.2 Completeness

The completeness of data in the Nordic cancer registries – and therefore also in NORDCAN – is quite high. However, it is not 100% - meaning that not absolutely all cases of cancer will be registered in these registries within a year. The completeness also differs from cancer site to cancer site, with lower completeness for non-symptomatic cancers of aggressive behavior and cancers where few diagnostics steps are taken (i.e. where the cancer registries have fewer sources of information) (2).

Of high importance is also the fact that although the completeness of cancer registration on case level is high, not all of these cancers are included in statistics like the one we provide in NORDCAN. Some cancer cases are, for instance, excluded due to international rules for multiple primary tumors (3). Even though the aggregate count in NORDCAN might be below 5 cases within a strata of age group, sex, region and year, this only reflects the number of cases being included in NORDCAN. The actual number of cases is often higher.

¹ If mid-year is not available, it can be calculated as the average of the current and the previous year.

4 NORDCAN risk assessment

4.1 Aggregate counts based on numbers in NORDCAN from 1943 to 2016

4.1.1 Smallest number of persons at risk

The smallest population, or *persons at risk*, for each strata of age group, year of diagnosis, sex and region for the period 1943 to 2016 (years currently in NORDCAN) is given in the table below.

Age group	Smallest population in NORDCAN
0-4	1633 Faroe Islands, women, 2010
5-9	1690 Faroe Islands, women, 1991
10-14	1550 Faroe Islands, women, 1996
15-19	1479 Faroe Islands, women, 1964
20-24	990 Faroe Islands, women, 1960
25-29	890 Faroe Islands, women, 1962
30-34	900 Faroe Islands, women, 1967
35-39	898 Faroe Islands, women, 1972
40-44	898 Faroe Islands, women, 1977
45-49	696 Greenland, women, 1968
50-54	595 Greenland, women, 1968

55-59	477 Greenland, men, 1968
60-64	352 Greenland, men, 1968
65-69	245 Greenland, men, 1968
70-74	149 Greenland, men, 1968
75-79	59 Greenland, men, 1968
80-84	12 Greenland, men, 1969
85+	7 Greenland, men, 1970

The smallest number of persons at risk in each stratum is quite large, except for in the two oldest age groups. The smallest numbers for most age groups appear more than 40 years back. Those that appear more recently contain a quite large number of persons at risk (more than 1500 people).

All the lowest aggregate counts for the oldest age group (85+) are from Greenland. The other country with low numbers is Faroe Islands, where 80 men in 1960 were from the oldest age group.

4.1.2 Smallest number of cases – smaller countries

The table below shows the five lowest total numbers of cancers within region, year of diagnosis and sex for each of the smaller countries in NORDCAN:

Faroe Islands	Greenland	Iceland
14 Faroe Islands, men, 1962	14 Greenland, men, 1968	45 Outside Reykjavik, women, 1977
15 Faroe Islands, men, 1963	19 Greenland, men, 1972	52 Outside Reykjavik, women, 1971
15 Faroe Islands, men, 1960	19 Greenland, men, 1970	52 Outside Reykjavik, women, 1980
15 Faroe Islands, men, 1961	22 Greenland, men, 1969	52 Outside Reykjavik, women, 1979
16 Faroe Islands, women, 1960	23 Greenland, men, 1973	53 Outside Reykjavik, women, 1976

These are total number of cases for all cancer sites combined. If age group and cancer site are taken into account, there will be strata (year of diagnosis, age group, sex, region, cancer site) containing only 1 case.

4.1.3 Smallest number of cases – larger countries

The table below shows the five lowest total numbers of cancers within region, year of diagnosis and sex for each of the larger countries in NORDCAN:

Denmark	Finland	Norway	Sweden
531 North Jutland, women, 1971	272 Oulu, women, 1958	199 Northern, women, 1957	714 Northern, women, 1970
546 North Jutland, women, 1974	277 Oulu, women, 1953	201 Northern, women, 1953	801 Northern, women, 1971
564 North Jutland, women, 1973	284 Oulu, women, 1954	205 Northern, women, 1955	825 Northern, women, 1973
567 North Jutland, women, 1977	288 Oulu, women, 1955	217 Northern, women, 1956	847 Northern, women, 1974
599 North Jutland, women, 1975	293 Oulu, women, 1956	218 Northern, women, 1959	860 Northern, women, 1972

4.2 Risks of disclosure of data concerning health through NORDCAN aggregate counts and statistics

#	Scenario	Possible consequence	C	L	Score	Justification	Acceptability	Action #
1	A data subject identifies him-/herself through NORDCAN aggregate counts	<p>Impression that others can also be identified through NORDCAN</p> <p>Loss of trust in NORDCAN</p>	1	4	4	There is a very high risk that a data subject can identify him- or herself through NORDCAN, given that the information is already known to this data subject. However, the data subject in question needs to be convinced that the necessary supplementary measures have been adopted, and that the risk of someone else identifying him/her through NORDCAN is negligible. See also risk scenario #2.	Acceptable	1
2	A data subject identifies one or more data subjects (other than him-/herself) through NORDCAN aggregate counts, either with or without other data available	<p>Unlawful disclosure of data concerning health</p> <p>Revealing information that a data subject is diagnosed with cancer</p> <p>Loss of trust in NORDCAN</p>	4	1	4	For a data subject to be identified with any certainty through aggregate counts in NORDCAN, the number of people who the data could possibly concern in a defined stratum needs to be insignificant. As shown in chapter 4.1.1, most strata currently contain more than 50 data subjects, leaving it virtually impossible to identify any data subjects. For the oldest age groups, the numbers are smaller, and the likelihood of identification is somewhat increased. However, the number of people at risk in the smallest strata is currently 7 in 1970, which is still too many to	Acceptable	2, 3, 4

						identify one or more data subjects without additional information.		
3a	The possibility of gaining more information about specific data subjects through NORDCAN aggregate counts for smaller countries (Faroe Islands, Greenland and Iceland)	<p>Unlawful disclosure of data concerning health</p> <p>Revealing information about what type of cancer a data subject is diagnosed with</p> <p>Loss of trust in NORDCAN</p>	3	2	6	<p>As shown in chapter 4.1.2, strata exist in NORDCAN where only 1 cancer case is diagnosed in each stratum of age group, sex, region and year of diagnosis and the total number of cancer cases are quite low.</p> <p>If already knowing that, for instance, your male neighbor in Greenland was diagnosed with cancer in 1968 at age 60, you could gain information on what kind of cancer or within which group of cancers he was diagnosed. However, if looking at risk scenario #2, you would already have to know everything else about your neighbor – including that he was diagnosed with cancer – to be able to get information on the specific diagnosis from NORDCAN.</p> <p>Since NORDCAN does not present all cases of cancers due to for instance multiple primary rules (see chapter 3.3.2), it would also be difficult to conclude with certainty that this information can actually be</p>	Attention	4, 5

						tied to the data subject you already have knowledge about.		
3b	A data subject gains more information about someone else through NORDCAN aggregate counts for larger countries (Denmark, Finland, Norway, Sweden)	Unlawful disclosure of data concerning health Revealing information about what kind of cancer a data subject is diagnosed with Loss of trust in NORDCAN	3	1	3	As we can see from the table in chapter 4.1.3, the total number of cancer cases for the larger countries included in NORDCAN are higher than for the smaller countries. This, in addition to the elements discussed in 3a, further decreases the risk of someone being able to gain more information about a data subject through aggregate counts and statistics in NORDCAN	Acceptable	4, 5

4.3 Disadvantages of not publishing NORDCAN aggregate counts and statistics

When assessing the risk of gaining more information about a data subject through in NORDCAN, the potential privacy risk of the individual need to be balanced against the interest of society. It would be considered both unethical and a breach of the purpose limitation principle in Article 5 of the GDPR to store data that cannot or is not used for the purpose it was collected. The main purpose of all cancer registries is contribute to the research, hereunder cancer epidemiology and cancer trends.

NORDCAN and similar statistical webtools provide an important source of information for health authorities, clinicians, researchers and the public. Freely available, high quality statistics such as NORDCAN is important for governments to plan the future impact of cancer patients in their countries or regions or to consider prevention measures to reduce the cancer burden. Clinicians can, for instance, use the data to compare incidence, mortality and survival between regions and countries to examine whether there are differences. For researchers, the data provide important epidemiological information on cancer trends and differences over the past half century (or more), which can be used to plan studies to better understand and reduce unintended differences. The general public can also benefit from these statistics. Not only can they help members of the public understand trends of their own cancer diagnosis or the diagnosis of their loved ones, but the available statistics may help the public understand the data governments need to use to plan for the future cancer burden.

However, to be truly useful for the above mentioned purposes, the aggregate data should be as detailed as necessary, while still preserving privacy and being anonymous. Decreasing the level of detail by, for instance, leaving out cancer sites or grouping multiple cancer sites into large

– and maybe heterogenous – groups, grouping several age groups together or grouping year of diagnosis (which is then also most often a proxy for year of treatment) to larger time periods can disguise trends and differences that need attention.

4.4 Conclusion of the risk assessment

Based on the above risk assessment (4.2) and the implemented risk-reducing measures already adopted (5.1) we conclude that the aggregate counts and statistics prepared by each cancer registry, using the R- and Stata-modules developed for this purpose, are rendered anonymous in such a manner that the data subjects are no longer identifiable. Accordingly, the data falls out of the scope of the GDPR and can be shared with IARC for use in the NORDCAN database and web tool.

5 Risk-reducing measures

5.1 Risk-reducing measures already implemented in NORDCAN

5.1.1 Risk reducing measures for aggregate counts and statistics in NORDCAN

Measure	Description	Comment
Individual data are stored in each country and not shared	Only aggregate counts are shared outside of the cancer registries	This reduces the risk of unauthorized disclosure.
Cancer sites are grouped in entities	NORDCAN contains a system of entities – groups of cancer sites – where cancer sites with a low number of cases are combined to make larger, more robust groups	The grouping of cancer sites to larger entities reduces the number of cells with rare/unique information, even when stratified for sex, age, geography and diagnostic period
The smallest geographical area is “region”	There are only two levels of geographical detail in NORDCAN: region and country. Countries of low population (Greenland, Faroe Islands) have no regions in NORDCAN, whereas the other countries have 2-6 regions. These regions are often used by the cancer registries themselves when publishing statistics	Grouping to larger levels of geographical areas increases the population at risk in each cell, reducing the risk of unlawful disclosure of data concerning health.
Only sex “male” and “female” are used	Although there is an option “other & unknown” for sex in NORDCAN, these cases are excluded from the dataset before it is transferred to IARC	Although cases of sex = other/unknown hardly exist in the Nordic countries, measures are made to make sure these are not included later. The main reason is that these data subjects would make up a quite small and probably easily recognizable group
5-year age groups are used up to 85+ (old version) or 90+ (proposed for new version).	Cancer statistics are shown in 5-year age groups (0-4, 5-9, 10-14 etc), not by individual age	Grouping in age groups increases the population at risk in each cell, reducing the risk of unlawful disclosure of data concerning health

The number of different variables in NORDCAN is small	The dataset behind NORDCAN is quite simple, and contains only a limitation of categories: age group, sex, geography, year of diagnosis and entity (cancer site/grouping of cancer sites), where entity is the only health related information. NORDCAN does not contain information on for instance stage or specific cancer type definitions	Few categories reduce the risk of anyone being able to use NORDCAN aggregate counts to identify or gain additional information about any data subject.
Same breakdown for all tables and graphs	In NORDCAN, the breakdown of the variables/categories are the same in most tables/graphs	Having the same breakdown in most tables and graphs reduces the risk of disclosure. Tables/graphs with a different breakdown have a lower degree of detail.
Transparency in data management and data flow	All processes in NORDCAN are transparent and possible to control by anyone. The R-packages for the NORDCAN-modules are shared at GitHub, and all documentation on call for data, data flow etc, including this risk assessment and the document on protecting confidentiality in NORDCAN, is available both at GitHub, at the NORDCAN website and at the ANCR website	Transparency in all processes reduces the perception of disclosure risk.

5.2 Action plan for risks identified through the risk assessment

The following table contains measures that can be taken either to reduce risk score or to ensure the score is kept on the current level.

Action #	#	Score	Possible vulnerability	Description of measure	Responsibility
1	1	4	Fear of being identified by others through NORDCAN Loss of trust in NORDCAN handling of privacy issues	The NORDCAN risk assessment will be made available through the NORDCAN website and upon request to the secretariat.	NORDCAN secretariat

				NORDCAN handling of data and means to protect the privacy of patients is transparent and well documented	
2	2	4	Revealing that a data subject is diagnosed with cancer	Although there is a wish from several different groups of NORDCAN users to split the oldest agegroup (85+) into 85-89 and 90+, we can keep 85+ as the oldest age group if number of people in the oldest age groups falls under 5 at risk for any strata	NORDCAN group
3	2	4	Revealing that a data subject is diagnosed with cancer	We should develop functionality in the NORDCAN R-modules which will alert the user if the persons at risk in a stratum falls below 5	NORDCAN development group NORDCAN secretariat
4	2	4	Revealing that a data subject is diagnosed with cancer	We might consider to group year of diagnosis into 5-year periods that can't be changed by the user	NORDCAN secretariat
	3a	6	Revealing information about what cancer a data subject is diagnosed with		IARC
	3b	3	Revealing information about what cancer a data subject is diagnosed with		
5	3a	6	Revealing information about what cancer a data subject is diagnosed with	We can present average annual number of cases within a 5 year period instead of true numbers for each year.	NORDCAN development group
	3b	3	Revealing information about what cancer a data subject is diagnosed with		NORDCAN secretariat

5.3 Sharing aggregate counts and statistics with IARC

The important risk-reducing measures for avoiding disclosure of data concerning health are given in chapter 5.1 and 5.2. However, we also need to make sure the counts and statistics sent to IARC keep their integrity – meaning that they are protected from unauthorized changes. These are the measures implemented to ensure integrity:

Measure	Description	Comment
Uniform naming of data files	The R-modules of NORDCAN ensure that all files transferred to IARC are named in a uniform way.	This reduces the risk of files going astray when they arrive in IARC.
Data zipped and password protected	All data will be zipped and password protected before they are transferred to IARC. The password will be transferred to IARC in a different channel than the zipped files.	This reduces the risk of data being accessible to others if the transfer should be intercepted.
Data transferred through a secure sFTP-channel	The zipped and password-protected files will be transferred to IARC through a secure FTP-channel	This reduces the possibility of the data transfer being intercepted by others.
Data stored in a protected database in IARC	All aggregate counts and statistics from the Nordic cancer registries are kept in a separate, protected database at IARC	This reduces the possibility of unauthorized persons getting access to NORDCAN data, and reduce the risk of data leakage or malicious access to data.
Only a few persons have access to data in IARC	NORDCAN-data is stored in a protected database with access control, and only a few clearly authorized persons have access to these data.	This ensures that the Nordic cancer registries have control of which persons have access to the data and how the data is used.

6 Literature

- (1) EU. **Article 4, GDPR - Definitions** <https://gdpr-info.eu/art-4-gdpr/>
- (2) Larsen IK., Småstuen, M., Johannesen, TB., Langmark, F., Parkin, DM., Bray, F., Møller, B. **Data Quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness.** European Journal of Cancer, Vol. 45, Issue 7, May 2009, Pages 1218-1231
- (3) IARC. **International rules for multiple primary cancers** http://www.iacr.com.fr/images/doc/MPrules_july2004.pdf
- (4) EU. **Article 5, GDPR – Principles relating to processing of personal data** <https://gdpr-info.eu/art-5-gdpr/>