2018-03-09

Nordic registry-based cohort studies : possibilities and pitfalls when combining Nordic registry data

Maret-Ouda, John; Tao, Wenjing; Wahlin, Karl; Lagergren, Jesper

Scandinavian Journal of Public Health. 2017 Jul;45(17_suppl):14-19.
http://doi.org/10.1177/1403494817702336
http://hdl.handle.net/10616/46265

Title: Nordic registry-based cohort studies – possibilities and pitfalls when combining Nordic registry data

Authors: John Maret-Ouda[1], Wenjing Tao[1], Karl Wahlin[1], Jesper Lagergren[1,2]

Affiliations:

[1] Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

[2] Division of Cancer Studies, King's College London, United Kingdom.

Correspondence and requests for reprints: Dr. John Maret-Ouda, Upper Gastrointestinal Surgery, Department of Molecular medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, 171 76 Stockholm, Sweden.

E-mail: John.Maret.Ouda@ki.se, Tel: +46 8 517 709 40, Fax: +46 8 517 762 80

Keywords: Register data; register-based; Scandinavia; population-based.

Abstract

Aims: All five Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) have nationwide registries with similar data structure and validity, as well as personal identity numbers enabling linkage between registries. These resources provide opportunities for medical research that is based on large registry-based cohort studies with long and complete follow-up. This review describes practical aspects, opportunities, and challenges encountered when setting up all-Nordic registry-based cohort studies.

Methods: Relevant articles describing registries often used for medical research in the Nordic countries were retrieved. Further, our experiences of conducting this type of study, including planning, acquiring permissions, data retrieval, and data cleaning and handling, and the possibilities and challenges we have encountered, are described.

Results: Combining data from the Nordic countries makes it possible to create large and powerful cohorts. The main challenges include obtaining all permissions within each country, usually in the local language, and to retrieve the data. These challenges emphasise the importance of having experienced collaborators within each country. Following the acquisition of data, data management requires the understanding of differences between the variables to be used in the various countries. A concern is the long time required between initiation and completion.

Conclusions: Nationwide Nordic registries can be combined into cohorts with high validity and statistical power, but the considerable expertise, workload, and time required to complete such cohorts should not be underestimated.

Introduction

Nationwide, administrative registries with long histories provide opportunities to conduct medical research on large cohorts with long time-periods and complete follow-up, as well as the possibility to study rare exposures and outcomes with sufficient statistical power. All of the five Nordic countries (Denmark, Finland, Iceland, Norway, and Sweden) have nationwide registries, containing virtually all individuals residing in those countries. The similarities between the registries in the Nordic countries make it possible to combine the registry data of each separate country into one larger cohort. Furthermore, all residents in the Nordic countries have personal identity numbers, allowing for linkages between different registries within each country. Hence, it is possible to retrieve large amounts of data pertaining to various aspects of each individual, such as factors relating to medical diagnoses and surgical procedures, sociodemographic characteristics and labour market participation. Multinational registry-based cohort studies are valuable in numerous medical research settings, including that they allow for the possibility of studying the effects and consequences of different treatments, interventions, and diseases, especially regarding rare outcomes requiring a long follow-up time, e.g. rare malignancies and late complications.

Planning

Conducting multinational registry-based cohort studies generally requires a team of researchers with special knowledge of the legislation and the relevant registries in each of the participating countries. It is valuable to collaborate with researchers with experience in registry-based studies and epidemiological and statistical methods when planning the study and retrieving the data needed. All contacts with the responsible authorities are best handled by the collaborators in that country. Following the creation of the research collaboration, a detailed study protocol should be written that outlines the studies to be conducted based on the cohort data. The protocol should define the registries to be included, inclusion and exclusion criteria for cohort participants, the exposures,

outcomes, and covariates of interest. Since coding systems vary between the country-specific registries (further elaborated below), a table containing all variables and codes of interest should be included. We recommend that the studies focus on key variables that are required for the validity of the studies, both with regards to the data management, and also due to legal aspects requiring that only data that are necessary for the study are retrieved. The variables are not always found in all registries, and the workload and experience required for retrieving and using each variable is substantial. The study protocol should include all collaborators, and we recommend personal meetings to agree upon a final study protocol, including a work and time plan and a co-authorship list before initiating the practical work on the project.

Registries

All Nordic countries share a similar structure in terms of most of their national registries, although they were initiated in different years (Figure 1). In this review we focus on registries that we have used in our all-Nordic studies and that might be particularly useful when conducting clinical cohort studies.

Patient registries: The basis for many clinical studies is the patient registries (hospital discharge registries) that generally include all inpatient care and often outpatient care as well. These registries contain data regarding codes for diagnoses and surgical procedures, admission and discharge dates from hospitals, and other information regarding hospital care, although these might vary between countries. The Nordic countries use different versions of the International Classification of Diseases (ICD) for the coding of diagnoses, and besides this, the early registration in Denmark used a national version of ICD-8 coding that needs to be retrieved locally [1]. Since the mid 1990's a Nordic collaboration has defined surgical procedures in the Nordic Medico-Statistical Committee (NOMESCO), making coding and selection of surgical procedures more homogenous between the countries [2]. The Danish National Patient Registry was founded in 1977, and reached complete

nationwide coverage in 1978 [3]. Initially it included only somatic inpatient care, but was later expanded to include both somatic and psychiatric in- and outpatient care [3]. The Danish National Patient Register has been validated in numerous studies. A recent systematic review of the available literature including 114 papers found the data validity, defined as positive predictive value, ranged from 15% to 100% depending on diagnosis [1]. The Finnish Hospital Discharge Register was founded in 1967 and it has been nationwide complete since its inception [4]. In 1994, this registry was replaced by the Finnish Care Register for Health Care, the main difference being that the Hospital Discharge Register included only data regarding inpatient care, while the Care Register for Health Care also contains data on specialised outpatient care and day-surgery [4, 5]. A recent review identified 32 studies validating the Finnish Hospital Discharge Register or the Finnish Care Register for Health Care, and found the positive predictive value to range from 75-99% for common discharge diagnoses [4]. The Icelandic Patient Registry has been centralised and nationwide since 1999, and contains continuously collected data from medical records at the hospitals [6]. The Norwegian Patient Registry was founded in 1997, initially containing only data on somatic health care, but since 2000 it has also contained psychiatric diseases [7]. Few validation studies have been conducted on the Norwegian Patient Registry to date. However, one study found a positive predictive value of 79.7% for stroke, and another study found that the number of fractures was overestimated by 19% in the registry [8, 9]. The Swedish Patient Registry was founded in 1964 and gradually included more regions to reach complete nationwide coverage in 1987 [10]. Only somatic inpatient care was included initially, but psychiatric care was added from 1973, and specialised outpatient care from 1997 onwards [10]. A review found that the positive predictive value was in the range of 85-95% for most diagnoses [10].

Cancer registries: There are well-established and national cancer registries in all Nordic countries that contain data on all malignant tumours. Besides the anatomical and histological classification of the tumours, these registries hold data on the date of diagnosis, basis of diagnosis (usually histopathology), and in some registries also tumour stage (TNM Classification of Malignant Tumours

staging). The Danish Cancer Registry was founded in 1942, and reporting to the registry became mandatory in 1987 [11]. The Finnish Cancer Registry was founded in 1953, and registration has been compulsory since 1961 [12]. The Icelandic Cancer Registry was founded in 1954, and registration of malignancies has been mandatory since its initiation [13]. The Norwegian Cancer Registry was founded in 1951, and registration has been compulsory since 1953 [14]. The completeness of the Norwegian Cancer Registry was 98.8% for the period 2001-2005, and the validity was 93.8% when compared to morphologic verification [14]. The Swedish Cancer Registry was established in 1958, and registration was mandatory from the beginning [15]. Approximately 98% of the cancers in the Swedish Cancer Registry are morphologically verified, and there was limited underreporting of cancers to the registry of approximately 3.7% based on an assessment of completeness in 1998 [16, 17]. The TNM registration has been validated in patients who have undergone surgery for oesophageal cancer in Sweden, concluding that the overall completeness of tumour stage was high (98.2%), although individual coding of each separate component (T, N, and M) needs to be improved [18].

Causes of death registries: The causes of death registries available in all Nordic countries contain data regarding date of death, main cause of death, and contributing causes of death. These registries also contain information on whether or not an autopsy was conducted. The Danish Causes of Death Registry has been electronically available since 1970, although a national registry has been kept since 1875 [19]. In Finland, death certificates are available from 1936 and a digital Causes of Death registry has been available since 1969 [20]. The Icelandic digital equivalent has been available since 1952, although causes of death have been published since 1911 [21]. The Norwegian Causes of Death Registry has been available since 1951 [22]. The digitally available Causes of Death Registry in Sweden was founded in 1961 [23].

Registries of the total population: The Nordic countries also maintain registries of the total population, containing information on birth dates, sex, immigration, emigration, education,

occupation, and socioeconomic status. In Denmark, there have previously been local population registries, however, since 1968 there has been a national registry of the total population [24]. The long history of population registers in Finland extends back to the 1530s, however, the registry has been centralised since 1969, and computerised since 1971 [25]. The Population Register of Iceland was founded in 1952 [26]. Statistics Norway has maintained national registries of the entire Norwegian population since it was founded in 1876, while personal identity numbers were introduced in 1964 [27]. Sweden started keeping population statistics in 1749, and in 1962 this was centralised under the Swedish Tax Agency, although the data handling is managed by Statistics Sweden [28]. A collaborative legislation makes it possible to also follow individuals moving between Nordic countries by means of the personal identity numbers and the registries of the total population.

Permissions

Ethical aspects when conducting registry-based research are of great importance, and although the data retrieved are not supposed to allow for identification of participating individuals, all permissions required in each country need to be obtained. The necessary permissions vary between the Nordic countries [29]. For registry-based cohort studies it is common practice that no personal identity numbers or other data that can be used to identify individuals are made available to the researchers. In all Nordic countries except for Denmark, there is a possibility to request that key codes are saved by the authorities, thus making renewed data retrieval and additional linkage of cohort members to other registries possible in the future. Keeping key codes usually requires special ethical approval with distinct arguments [29]. For registry-based research as described above, no ethical permission is needed in Denmark or Finland for the data acquisition, providing all data are from registries only. However, approval is needed from the Danish Data Protection Agency, and in Finland approvals are needed from the National Institute for Health and Welfare, Statistics Finland,

and the Population Registry Centre. The Icelandic regulations require ethical permission from the National Bioethics Committee, as well as permission from the Data Protection Authority. In Norway, ethical permissions are required from the regional ethical committee. In Sweden, ethical permission from the relevant regional ethical committee is required if sensitive data are handled, and besides this, approvals from the relevant governmental registry holder (Statistics Sweden and the National Board of Health and Welfare) are required [30, 31]. The intent of merging the data for multiple countries needs to be highlighted in the applications, although this does not affect the retrieval of permissions in general. Applying for, and acquiring the necessary permissions is generally one of the most time-consuming aspects of this type of study. It should be stressed that permissions required from the registry holders are not based on another ethics assessment, but these are separate assessments with a focus on the laws that regulate the secrecy and integrity of the individuals included in the registries. At least one year should be allocated in the planning process for the retrieval of all necessary permits. In general, the approval processes are preferably managed in close collaboration with researchers in each country, since the necessary documents are generally handled in the local language.

Data retrieval and management

Once the permissions have been obtained, the data are retrieved from the various authorities in each country. The data used for clinical cohort studies are typically selected based on a predefined exposure, such as a disease or a surgical procedure, or a combination of both. Following the identification of individuals in each of the countries meeting the inclusion criteria for the cohort, other medical history of the cohort members is often retrieved from the patient registries to evaluate comorbidities. The registry linkages are conducted at the agencies that maintain the registries. All individual data that can identify the individual, such as personal identity numbers, are then removed and the data are pseudo-anonymised through replacement of this information with

an arbitrary number. This arbitrary number is, however, the same for each individual in all registries within the country, enabling further linkages between registries. The data are generally delivered on discs or through safe internet connections for further storage on safe and well-protected servers within the university networks. Due to regulations, Danish data are not permitted to be handled outside the country, which is not the case in the other Nordic countries. To overcome this issue in all-Nordic studies, data from the other Nordic countries are retrieved and stored on servers maintained by Statistics Denmark. The data are accessed by one or a few named researchers through an individual secure virtual private network (VPN) connection for data management and statistical analyses. Since no data are allowed to leave Denmark, it is not possible to conduct any data handling or management in any other country.

Once delivered, the data need to be cleaned and merged. The first aspect is to identify each variable that will be used for the purposes of the study in each of the registries, and thereby finding the corresponding variable in registries from the other countries. Following this, the variables need to be transformed into the same format, making merging of the data sets possible. The datasets from each registry are delivered as separate files that can by merged by means of the pseudo-anonymised arbitrary number.

Opportunities and challenges

Combining the national registries from the Nordic countries increases the statistical power of the studies. This makes it possible to study rare exposures as well as rare outcomes with long follow-up, and possibilities to find small differences between exposure groups. The diverse spectrum of information available from the registries for each individual makes it possible to adjust for potential confounders in the statistical analysis. For example, registration of all diagnoses and surgical procedures enables adjustments for comorbidity. The nationwide property of the registries reduces the risk of selection bias, and losses to follow-up are minimised. Furthermore, registry-based

research in its design makes it possible to study research questions where other study designs, such as randomised clinical trials, would be unethical or impossible to conduct.

There are also several challenges when conducting a multinational registry-based cohort study. One of the main challenges is the time-consuming and complex bureaucracy associated with creating these cohorts, specifically obtaining all relevant permits and retrieving all data. Much of this is done in the local language, making a local collaborator invaluable. Data management is another major and time consuming concern of a multinational study, which requires knowledge of the variables of the registries as well as statistical competence. There is also a risk of differences regarding the quality of data between the countries, mainly since different countries have varying routines regarding coding as well as quality control of the registries. We therefore recommend that sufficient time is allocated for visual examination and tabulation of the variables, and that this should be done by both a statistician and clinician, to ensure the correctness and reliability of the data. There might also be substantial discrepancies between the registries, and we therefore recommend thorough comparisons of the variables (such as visual examination, tabulation, and hypothesis testing) before conducting the statistical analyses. Multilevel models should be considered in order to deal with possible clustering within each country. Potential differences in data between countries may have various explanations. Underlying reasons for differences include different clinical praxis, differences regarding coding of diagnoses and interventions among clinicians, and different administrative strategies in coding and data management. Further, it cannot be ruled out that economic incentives and political regulations might influence coding between countries, especially regarding completeness of chronic co-morbidities that are not the main reason for patients seeking healthcare. Furthermore, since personal identity numbers in the different registers were not standardised until 2007, the Norwegian registers can be linked only from this year onwards. Also, administrative registries lack information on e.g. lifestyle factors and health-related quality of life, and some diagnoses are associated with a higher risk of misclassification, e.g. obesity, which is seldom the

main reason for patients seeking health care. Besides this, there are also local differences in coding due to varying clinical praxis between the Nordic countries.

Conclusions

The nationwide registries available in the Nordic countries make it possible to create large multinational cohorts by combining data from several registries of similar design and contents. This makes it possible to assess rare exposures as well as rare outcomes with sufficient power. Key recommendations that we want to highlight based on our experience of all-Nordic cohort studies include to complete a clear study protocol before initiating the study, involve researchers and statisticians with experience from registry-based research from each country, and allocate sufficient time and expertise, including experienced biostatisticians and epidemiologists, to clean, merge and analyse all data. Before the data are retrieved and managed, differences between the registries and clinical praxis within the countries should be carefully considered. Further, only include key variables and define the needed variables in each of the registries in all countries. The data are ready for statistical analysis only after extensive data management.

Funding

This article was funded by the Swedish Research Council (Grant no. 340-2013-5478). The funding source had no role in the design, conduct, analysis, or reporting of the study.

References

[1] Schmidt M, Schmidt SA, Sandegaard JL, Ehrenstein V, Pedersen L and Sorensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clin Epidemiol*. 2015; 7: 449-90.

[2] Nordic Welfare dataBASE. *NOWBASE*, http://nowbase.org/da/current-projects (2016, accessed Aug 9 2016).

[3] Lynge E, Sandegaard JL and Rebolj M. The Danish National Patient Register. *Scand J Public Health*. 2011; 39: 30-3.

[4] Sund R. Quality of the Finnish Hospital Discharge Register: a systematic review. *Scand J Public Health*. 2012; 40: 505-15.

[5] National Institute for Health and Welfare. *Care Register for Health Care*, https://www.thl.fi/fi/web/thlfi-en/statistics/information-on-statistics/register-descriptions/care-register-for-health-care (2016, accessed Aug 2 2016).

[6] Directorate of Health. *Vistunarskrá heilbrigðisstofnana*, http://www.landlaeknir.is/tolfraedi-og-rannsoknir/gagnasofn/gagnasafn/item12464/Vistunarskra-heilbrigdisstofnana (2016, accessed Aug 9 2016).

[7] The Norwegian Directorate of Health. *Om Norsk Pasientregister*, https://helsedirektoratet.no/norsk-pasientregister-npr/om-npr (2016, accessed Aug 2 2016).

[8] Varmdal T, Bakken IJ, Janszky I, et al. Comparison of the validity of stroke diagnoses in a medical quality register and an administrative health register. *Scand J Public Health*. 2016; 44: 143-9.

[9] Lofthus CM, Cappelen I, Osnes EK, et al. Local and national electronic databases in Norway demonstrate a varying degree of validity. *J Clin Epidemiol*. 2005; 58: 280-5.

[10] Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health*. 2011; 11: 450.

[11] Gjerstorff ML. The Danish Cancer Registry. *Scand J Public Health*. 2011; 39: 42-5.

[12] Finnish Cancer Registry. *Finnish Cancer Registry: Forms and Instructions*, http://www.cancer.fi/syoparekisteri/en/registration/forms-and-instructions/ (2016, accessed Aug 5 2016).

[13] The Icelandic Cancer Society. *About the Icelandic Cancer Registry*, http://www.krabbameinsskra.is/indexen.jsp?id=aboutics (2016, accessed Aug 3 2016).

[14] Larsen IK, Smastuen M, Johannesen TB, et al. Data quality at the Cancer Registry of Norway: an overview of comparability, completeness, validity and timeliness. *Eur J Cancer*. 2009; 45: 1218-31.

[15] The National Board of Health and Welfare. *The Swedish Cancer Registry*, http://www.socialstyrelsen.se/register/halsodataregister/cancerregistret/inenglish (2016, accessed Aug 3 2016).

[16] The National Board of Health and Welfare. *Cancer Incidence in Sweden 2011*, http://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/18919/2012-12-19.pdf (2012, accessed Aug 8 2016).

[17] Barlow L, Westergren K, Holmberg L and Talback M. The completeness of the Swedish Cancer Register: a sample survey for year 1998. *Acta Oncol*. 2009; 48: 27-33.

[18] Brusselaers N, Vall A, Mattsson F and Lagergren J. Tumour staging of oesophageal cancer in the Swedish Cancer Registry: A nationwide validation study. *Acta Oncol*. 2015; 54: 903-8.

[19] Helweg-Larsen K. The Danish Register of Causes of Death. *Scand J Public Health*. 2011; 39: 26-9.

[20] Statistics Finland. *Causes of Death*, http://www.tilastokeskus.fi/meta/til/ksyyt_en.html (2016, accessed Aug 5 2016).

[21] Statistics Iceland. *Deaths*, http://www.statice.is/publications/metadata?fileId=19603 (2016, accessed Aug 5 2016).

[22] Norwegian Insitute of Public Health. *Dødsårsaksregisteret*, https://www.fhi.no/hn/helseregistre-og-biobanker/dodsarsaksregisteret/ (2016, accessed Aug 2 2016).

[23] The National Board of Health and Welfare. *Dödsorsaksregistret*, http://www.socialstyrelsen.se/register/dodsorsaksregistret (2016, accessed Aug 3 2016).

[24] The Civil Registration System. *The National Register (Folkeregistret): Historie*, https://cpr.dk/cpr-systemet/historie/ (2016, accessed Aug 12 2016).

[25] Population Register Centre. *History*, http://vrk.fi/en/history (2016, accessed Aug 15 2016).

[26] Registers Iceland. *Population register*, http://www.skra.is/english/english/ (2016, accessed Aug 15 2016).

[27] Statistics Norway. *About us*, http://www.ssb.no/en/omssb/om-oss (2016, accessed Aug 15 2016).

[28] Statistics Sweden. *Statistics Sweden's History*, http://www.scb.se/statistics-swedens-history (2016, accessed Aug 15 2016).

[29] Ludvigsson JF, Haberg SE, Knudsen GP, et al. Ethical aspects of registry-based research in the Nordic countries. *Clin Epidemiol*. 2015; 7: 491-508.

[30] Statistics Sweden. *Beställa mikrodata [Swedish]*, http://www.scb.se/sv_/Vara-tjanster/Bestalla-mikrodata/ (2016, accessed Aug 15 2016).

[31] The National Board of Health and Welfare. *Beställa data/statistik [Swedish]*, http://www.socialstyrelsen.se/register/bestalladatastatistik (2016, accessed Aug 15 2016).

(See separate file ”Fig 1.eps”)

Figure 1. Timeline showing the year of complete coverage of the patient registries, compulsory reporting to the cancer registries, and electronically available causes of death registries in the Nordic countries.