Scandinavian Journal of Public Health, 2017; 45(Suppl 17): 5–13

ORIGINAL ARTICLE

Estimating incidence and prevalence from population registers: example from myocardial infarction

KARIN MODIG[1], ANITA BERGLUND[1], MATS TALBÄCK[1], RICKARD LJUNG[1] & ANDERS AHLBOM[1,2]

[1]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, and [2]Centre for Occupational and Environmental Medicine, Stockholm County Council, Sweden

Abstract

Aim: To illustrate how the fundamental epidemiological measures, incidence rate and prevalence proportion, can be estimated based on Swedish population registers using acute myocardial infarction (MI) as an example, together with a discussion about the analytical decisions. *Methods:* All individuals in Sweden aged 60–89 (born 1904–1954) during the study period 1994–2014 were identified through the Total Population Register. Cases of MI were defined and identified from information on hospital admissions and causes of death. Incidence rates of all, first, and recurrent MI were calculated together with prevalence proportions. *Results:* The incidence rate of all, first, and recurrent MI declined over the study period. While the incidence rates of first MI are lower for women than men, the incidence rates of recurrent MI are considerably higher but similar for men and women. The prevalence calculated with duration of disease set at 28 days also declined. This was despite improved survival from MI and increased life expectancy over the same period meaning that the decline in incidence was large enough to compensate for increased survival. ***Conclusions:*** **Calculating incidence and prevalence of diseases using population registers requires detailed and well-reasoned definitions. The definitions will affect both the study population and the number of disease events and it is essential that the cases and the study population are defined in a coherent way. Different measures of disease occurrence contribute with different aspects of the disease panorama and a joint interpretation contributes to a thorough understanding of the disease development in a population.**

Key Words: *Epidemiology, epidemiologic measurements, population registers, population at risk, incidence, prevalence, recurrence, cause of death, cardiovascular diseases*

Introduction

In order to plan health care resources for the future it is essential to monitor and follow disease rates in the population. It is also necessary in order to plan implementations and actions aimed at improving health, and for evaluating such actions. Continuous monitoring of disease trends also makes it possible to observe sudden or unexpected changes in disease risks, perhaps attributed to a new risk factor. Incidence and prevalence are core indicators of public health and are used for the purposes described above, and for calculations of disease burden [1]. In addition, incidence and prevalence constitute necessary input in simulation models designed to make projections of future population health [2]. Many countries lack nationwide data about diseases why incidence and prevalence have to be estimated and modelled based on information for only a selected part of the population [3]. The Nordic countries, however, have comprehensive nationwide registers that contain health related data of the whole

Correspondence: Karin Modig, Unit of Epidemiology, Institute of Environmental Medicine, Box 210, Karolinska Institutet, SE-171 77 Stockholm, Sweden. E-mail: karin.modig@ki.se

Accepted 29 January 2017

DOI: 10.1177/1403494817702327
journals.sagepub.com/home/sjp

population [4], such as hospital admissions. These data are a great source of information for epidemiological research and are for some diseases used as the basis for estimation of incidence and prevalence. Even if hospital-based data have the disadvantage that they include only those cases that lead to hospitalization, certain diseases, e.g., acute myocardial infarction (MI), are well suited to study based on registers of hospital admissions since almost all events result in a hospital admission (or death). Cases of MI can thus be identified by combining information about hospital admissions and death [5]. Yet, discrepancies from the true number may still occur due to several reasons, such as (i) all MI patients do not seek hospital care, (ii) some present with atypical symptoms, (iii) the diagnostic criteria that are used at hospitals do not have 100% sensitivity and specificity, (iv) on some instances the diagnosis that is made does not meet the criteria, and (v) coding errors occur in the medical records. However, despite these limitations, studies have shown that hospital admissions and causes of death can be combined to cases of MI with reasonable accuracy [6].

In addition to identifying the cases accurately, epidemiological studies also require information about the size of the relevant population, i.e., that generates the cases. In particular the calculation of incidence rates requires access to the population at risk, i.e., excluding prevalent cases, or recurrent cases if the aim is to estimate the incidence of the first disease event. Calculation of the prevalence proportion requires information about the total population at a specific point in time (both the nominator and the denominator). The population registers in the Nordic countries offer a unique opportunity of this, but nevertheless require detailed and well-reasoned definitions.

The aim of this paper is to illustrate how the fundamental measures of disease occurrence, incidence rate and prevalence proportion, can be estimated based on population registers. MI will be used as an example and incidence rate and prevalence proportion of MI will be calculated, together with a discussion about the analytical decisions. Possibilities and limitations in relation to extensions to other diseases will be discussed.

Material and methods

Material

We draw on population register information for all individuals born 1904–1954 and living in Sweden from January 1, 1987 and followed until December 31, 2014 [7]. Results are presented for ages between 60 and 89 years old. Individual data were used.

The population was identified from the Total Population Register. Cases of MI were identified by combining hospital admissions or causes of death through the National Patient Register and the Cause of Death Register. The International Statistical Classification of Diseases and Related Health Problems (ICD) codes were used to identify MI admissions or deaths with codes I21 and I22 in ICD-10 or 410 in ICD-9. MI was defined based on the main or secondary diagnoses of inpatient care in the National Patient Register or from the underlying or contributing causes of death in the Cause of Death Register.

Methods

Incidence rate, recurrent events, and prevalence proportion. Age-specific incidence rates were calculated for men and women separately for first, recurrent, and all MI for each year between 1994 and 2014 when the individuals were between 60 and 89 years old. The number of cases for each age (attained age) and calendar year was divided by the corresponding number of person years at risk (the denominator). In order to estimate the first disease one would in theory need information about the entire disease history of individuals. When using register data there is a time point when the register starts resulting in left truncation, i.e., before a certain time there is no information about the disease history of individuals, meaning that it is not possible to definitely define a first occurrence of disease. A practical way to handle this is to apply a period at start where events are disregarded in calculations of disease occurrence and only individuals free of disease after the wash out period are followed up. A 7-year period was chosen for all individuals to ensure the same likelihood to capture first events of MI, which is the same wash out period used by the National Board of Health and Welfare [8]. The National Patient Register started to have national coverage in 1987 meaning that the follow up in this study is presented from 1994 and onwards. When estimating all MI no such wash out period is needed but for coherence, results are presented from 1994 as well. In addition to apply the 7-year period at start of follow up for everyone, we applied two procedures for calculating the incidence of first MI. In the first scenario the 7-year period was applied continuously over the follow up period, meaning that individuals were allowed to return to the population under risk of a "first" MI 7 years after the previous event. By this procedure the time series was set to be comparable over time. In the other scenario, individuals could only experience one first MI, regardless of how many years that passed by since the

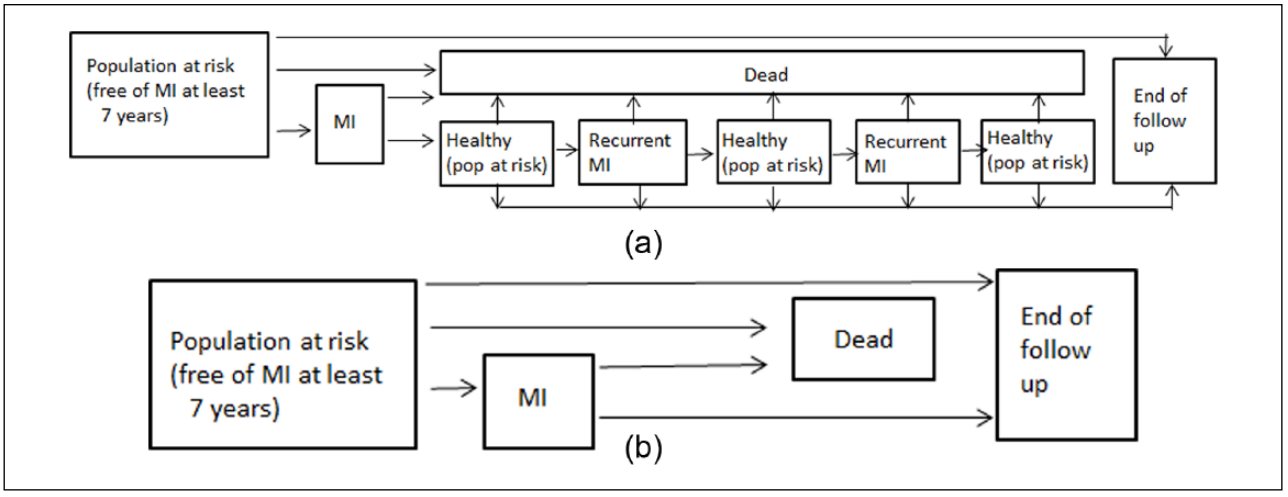

Figure 1. Flow chart of population at risk and MI events when MI is defined as (a) a non-chronic disease and (b) as a chronic disease.

previous event (a period that became longer as the follow up time increased).

The incidence rate of all MI is all cases of MI in relation to the total number of person years at risk, i.e., all person years in the study population. A recurrent MI is defined as any MI that is not the first (based on the 7-year definition of a first MI) and occurs at least 28 days apart from the onset of the previous event [9]. This is because the World Health Organization and the Swedish National Board of Health and Welfare have been using 28 days to separate events resulting from an ongoing MI from those related to a new MI. Another hospital admission for MI, or death, within the 28 days is considered as related to the ongoing MI and does not represent a new MI. The population at risk of recurrent MI thus consists of everyone who had a previous MI within the past 7 years and are not currently (28 days) experiencing an MI. The incidence for all MI is the average of the incidence rate for first MI and recurrent MI weighted according to the proportion in the population who are at risk of first and recurrent MI respectively. That is, the weights are determined by the prevalence of previous MI in the population.

Calculations of MI prevalence are not common, most investigations only consider incidence. This is because MI is usually seen as a non-chronic disease. Mortality is also high during the first period following the onset but declines quickly with time. However, a first MI can be seen as a first manifestation of lifelong coronary artery disease [10], and, thus, identifies a group of people at raised risk for heart disease. Therefore MI can be regarded as a chronic disease and the duration of first MI thus lifelong (see Figure 1(b)). We therefore calculated the prevalence proportion of MI as the proportion in the population having ever experienced an MI. However, because of the need for a wash out period at start we had to apply the 7-year period continuously over the period. Otherwise the accumulation of lifetime prevalent cases would be 7 years at the start of the period but 27 years at the end of follow up, making the comparison unjust. In addition we calculated the prevalence proportion with duration of 28 days.

For both all MI and first MI we calculated how many prevalent cases of MI there was at the end of each month and divided this with the number in the population at the end of the month. A yearly average of the monthly figures was calculated. There is a general relationship between incidence rate, prevalence proportion, and duration of disease so that the prevalence equals the incidence times the duration; the calculation of the duration must consider the mortality from the disease, which, e.g., reduces the duration from 28 days to about 20.

Both Figure 1(a) and (b) display the interplay between incidence, fatal cases, disease duration, and prevalence. In particular, it indicates how decreased incidence drags the prevalence downwards and how improved survival with the disease (leading to increased duration) pushes the prevalence upwards.

An ethics approval for this study was obtained from the regional ethics committee in Stockholm, Dnr. 2011/136-31/5.

Results

The results are based on 19,216,124 person years for men and 22,896,908 person years for women. In total, 355,698 events of MI occurred among men and 249,392 among women.

Incidence of first-and all MI declined over the study period (Figure 2(a) and (b)). Only among the oldest age group was there an increase in the beginning of the period but after around year 2002 a prominent decline. The patterns are very similar for men and women. For first MI, the solid black line presents the results when a 7-year wash out period was used continuously over the period and the dashed black line when a 7-year disease free period was applied only at the start of the period. Results are similar even if the curves diverge in the oldest ages where most cases occur.

Figure 3 displays the trend of recurrent events calculated with consideration of the seven year cut off for definition of first MI. Overall, the incidence of recurrent MI has halved over the study period for both men and women (Figure 3). The decline is most pronounced in the highest age groups. In the oldest age group, there was however a large increase in the beginning of the period. The incidence rates of recurrent MI are almost equal between men and women.

Figure 4 presents the prevalence proportion of all MI, based on the assumption that the duration of a non-fatal MI is 28 days. The prevalence came down, in general, over the study period and estimates to around 0.7 per 1000 for men and 0.4 per 1000 for women in 2014. That is, among those between 60 and 89 on average about 1 man out of 1000 and 1 woman out of 2000 had a MI during the previous month. The exception is the three oldest age groups where the prevalence increased in the beginning of the period to come down around year 2002 and onwards but more pronounced compare to the younger age groups.

The prevalence proportion defined as having had an MI during the past 7 years also decreased, from around 5.5% to 4.5% among men and from 2.5% to 2% among women (average over all ages 60–89 years) (Figure 5).

Discussion

In this paper we have used Swedish population registers to assess the disease burden of MI as an example. A number of aspects have been illustrated. First, the issue of left truncation of data, that is; in order to capture the first disease event when the entire disease history is not known a wash out period has to be applied. Second, we illustrated that the wash out period can be used continuously onward in order to create comparable time periods, or if not, as time passes by the accuracy of a first disease event will be closer to reality – but less comparable with the values at the start of the follow up period. Which procedure is more correct? It depends on the study purpose. If one accurately wants to describe first events of MI in the population as long wash out period as possible is desirable. The draw back with this is that in the beginning of the period the incidence rate of first MI will be somewhat overestimated as compared to the later period and the decrease in incidence rate may thus appear more pronounced than what is real. The magnitude of this effect is visible in Figure 2. In the oldest age groups, the difference is about 10% between rates calculated with the 7-year wash out and rates calculated without the wash out. The effect is only seen in the older age groups because these are the only ones old enough to have experienced an MI prior to the observation period.

A third illustration was the estimation of recurrent events which requires a definition of the duration of the disease, that is; after what time period should another disease event count as a new event and not related to the first, and when should individuals be considered to be at risk for a new event. It is also important that the denominator (the population at risk) consists of individuals having experienced a previous disease event and therefore is at risk of recurrence. In order to keep comparability over time, the 7-year period was applied also for recurrent events. That is, after a 7-year disease free period individuals were no longer under risk for a recurrent event but returned to being at risk of a first event. An alternative approach to estimating the incidence rate of recurrent events could be to look at the time-to-new event with Kaplan Meier-curves for each new yearly MI cohort. However, in this case when estimating incidence of first and all MI, it is of value to estimate incidence of recurrent events in a similar manner. Finally, we calculated the prevalence proportion of MI and illustrated that the choice of definition of the duration will result in different number of prevalent cases.

Interpretation of our findings

We found a decline in the incidence rate of all-, first-, and recurrent MI over time, in line with other studies [11–13]. The age specific incidence rate of first MI was expected to decrease given previous research [12]. For recurrent MI, however, a decline was not evident given that the first MI has been postponed to higher ages and the survival of the first MI has improved, both factors potentially making individuals more fragile after their first MI and therefore at higher risk of recurrent events. This turned out not to be the case. The risk of a recurrent MI is of course much higher than the risk of a first MI, but the risk has decreased considerable over time. It is beyond the

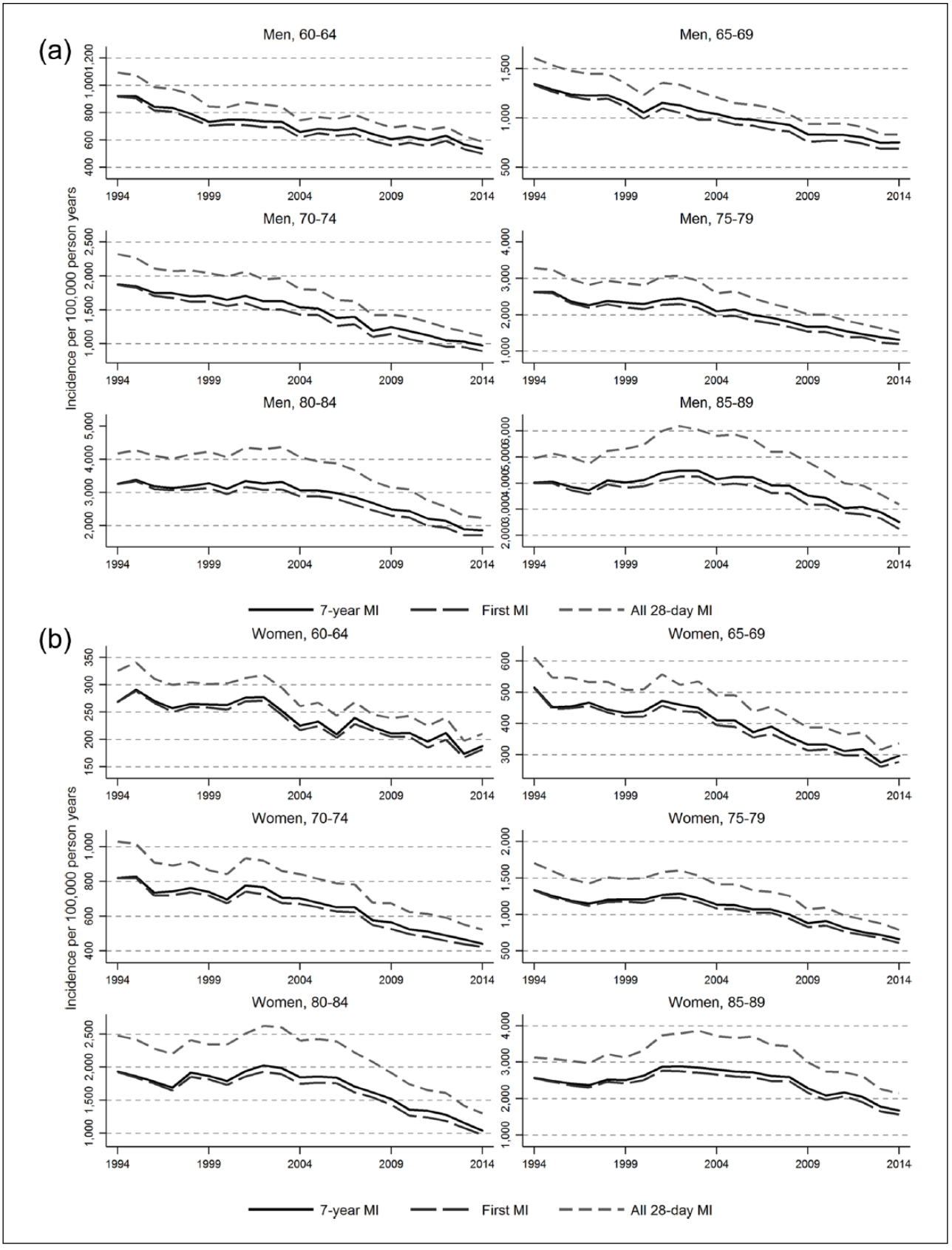

Figure 2. Incidence rate of first and all MI over time, 1994-2014: (a) men; (b) women.

Figure 3. Incidence rate of a recurrent event of MI over time, 1994–2014: men and women.

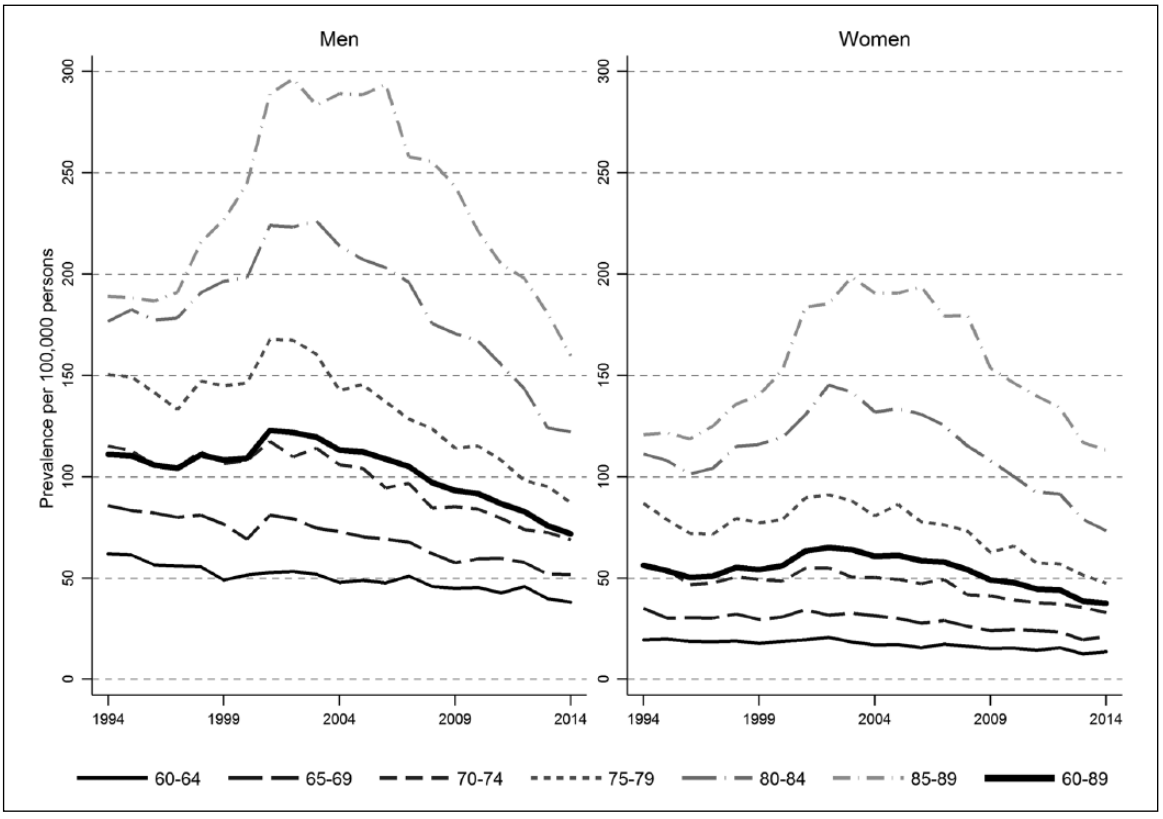

Figure 4. Prevalence proportion of MI between 1994 and 2014, 28 days duration: men and women.

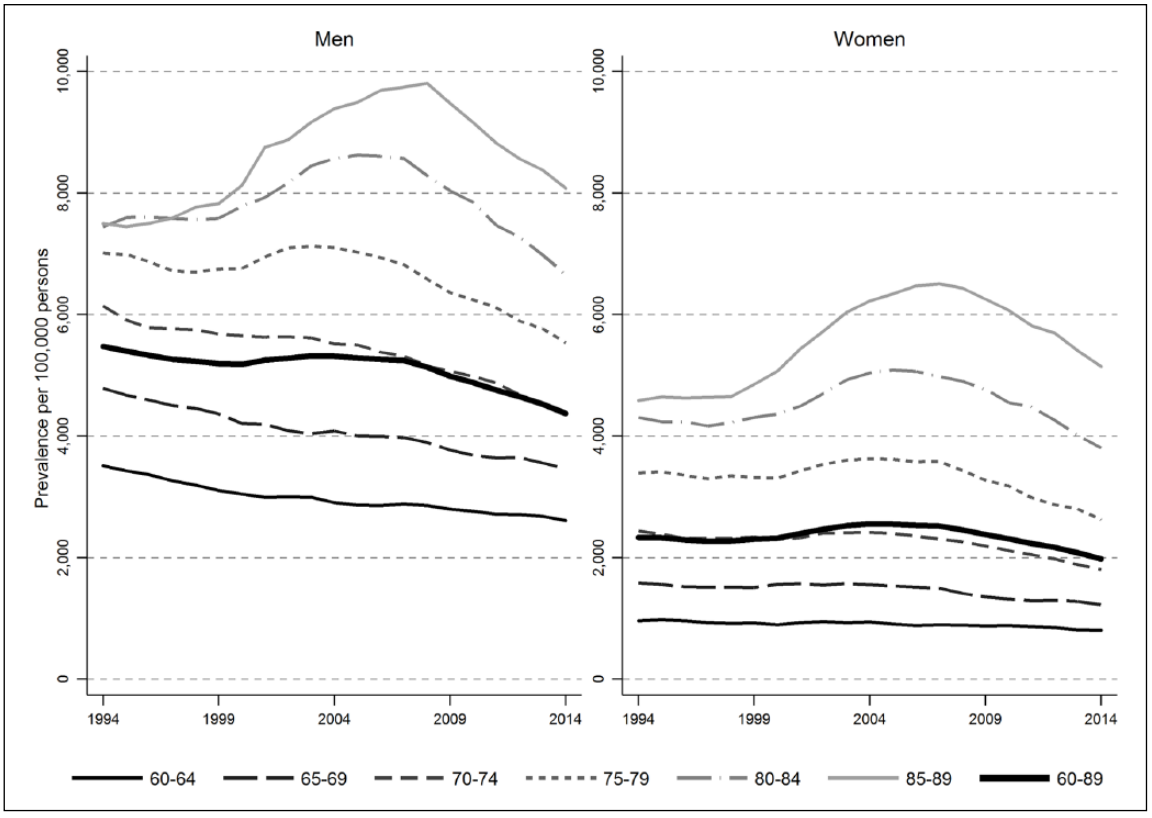

Figure 5. Prevalence proportion of MI between 1994 and 2014, 7-year duration: men and women.

scope of this paper to explore the mechanisms behind this decrease but it is likely that secondary prevention and treatment play roles. Worth mentioning is that the incidence rate of a recurrent MI is as large for women as for men even if women have about half the risk of a first MI as compared to men.

The prevalence proportion of MI was first calculated based on the definition that the duration of a non-fatal MI is 28 days. It declined over time, slightly more so for men than for women. Overall, the decline in the prevalence proportion is much smaller than the decline in the incidence rate, revealing the improvements in survival occurring over the study period [12]. Since there is a general relationship between incidence rate, prevalence proportion, and duration of disease these figures can be compared to the incidence rate results and show the mean duration of disease to be in the order of 20 days. Case fatality explains why the mean duration is 20 days rather than 28.

The survival from MI has increased considerable in parallel with the overall survival of the population, meaning that not only do individuals survive their MI to a higher extent; they also live longer in general. Both factors increase the prevalence proportion. Therefore, the observed decline in the prevalence proportion must be driven by the decline in the incidence rates of MI. The decline in the prevalence also indicates that the decline in incidence was large enough to "compensate" for increased survival. When the duration of MI was set to seven years, there was still a decline in the prevalence proportion for men but the proportion remained fairly stable for women. Knowing that the incidence rates have declined substantially, the stability in the 7-year prevalence must stem from an improvement in survival, either short term survival (case fatality) or improvement in overall survival causing the duration of the disease to increase.

The finding of a decline in the prevalence proportion is in line with a Danish study that found the MI prevalence between 2000 and 2009 to decline. In 2009 the prevalence proportion was around 3% for men and 1% for women in ages 35 years and older [14], lower than what we found. However, since we have a much older population it is expected that our prevalence proportion is higher. Similar findings was seen in a study from Canada where the prevalence proportion of MI in ages 20 and above was around 2% in 2004 estimated as the proportion having experienced at least one MI between 1988 and 2004 [15]. Figures were in both studies presented as an average

over all ages, and as we have shown, for both incidence and prevalence, the results differ between ages.

Use of register data for measuring incidence and prevalence

MI is particularly well suited for studies on register-based data because of the nature of the disease and because of standardized diagnostic criteria. There are a few other diseases equally suited, e.g., stroke and hip fracture [16]. However, diseases without a clear onset, like diabetes and rheumatoid arthritis, often do not require hospital inpatient care at disease onset, hence hospital admission data are not suited for studying incidence rate of such diseases. For some diseases where specific medications are prescribed to all or nearly all patients, e.g., diabetes, the Swedish Prescribed Drug Register can be used [17,18]. For cancer there is a specific disease register, the Cancer Register, which includes all diagnosed primary cases of cancer in Sweden. In addition to nationwide health data registers where reporting is compulsory by law, Sweden has around 100 quality registers [19]. These registers are not mandatory for health care providers to report to, but for some diseases the completeness is high and they can thus be used for estimations of disease occurrence.

Changes in diagnostic criteria and improvement in diagnostic tools will also influence the disease occurrence as measured by reporting to registers. For example, new diagnostic criteria for MI increased the incidence from one year to another in 2001. Also, the introduction of high-sensitivity cardiac troponin T assessment will detect more MIs. Further, the sensitivity and specificity of reported diagnoses to the registers need to be assessed. For example, reported diagnoses for the very old could be of lower quality due to less diagnostic accuracy and lower autopsy rates. All of these aspects need to be taken into consideration when interpreting the findings.

Conclusions

Calculating incidence and prevalence of diseases using population registers require detailed and well-reasoned definitions. The definitions will affect both the study population and the number of disease events and it is essential that the cases and the study population are defined in a coherent way. This may have implications for interpreting how health has developed over time in the population and it requires caution when comparing results from different studies and countries since the procedure may be different. If the total population serves as the population at risk without considering that ongoing cases are not at risk for the disease, the incidence rate will appear too small. Although the magnitude of this issue depends on how common the disease is. The different measures of disease occurrence, incidence, prevalence, recurrent events and survival, all contribute with different aspects of the disease panorama and interpreting them together contributes to a more thorough understanding of the disease development in a population than if only one of them is presented.

The population registers in the Nordic countries are unique and provide a wealth of data of immense importance to science and to planning and evaluation of health care and public health interventions but they must be used with skill and insight.

Declaration of conflicting interests

The authors declare that there are no conflicts of interest.

Funding

This research was funded by FORTE, grant number 2016-00685. Forte is a research council funding research on health, working life and welfare.

References

[1] Lopez AD, Mathers CD, Ezzati M, et al. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* 2006;367:1747–57.
[2] Hoogenveen RT, van Baal PH and Boshuizen HC. Chronic disease projections in heterogeneous ageing populations: approximating multi-state models of joint distributions by modelling marginal distributions. *Math Med Biol* 2010;27:1–19.
[3] Giampaoli S, Palmieri L, Capocaccia R, et al. (2001) Estimating population-based incidence and prevalence of major coronary events. *Int J Epidemiol* 2001;30(Suppl 1):S5–10.
[4] Rosen M. National Health Data Registers: a Nordic heritage to public health. *Scand J Public Health* 2002;30:81–5.
[5] Ahlbom A. Acute myocardial infarction in Stockholm: a medical information system as an epidemiological tool. *Int J Epidemiol* 1978;7:271–6.
[6] Ludvigsson JF, Andersson E, Ekbom A, et al. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;11:450.
[7] Ludvigsson JF, Almqvist C, Bonamy AK, et al. Registers of the Swedish total population and their use in medical research. *Eur J Epidemiol* 2016;31:125–36.
[8] Swedish National Board of Health and Welfare. Myocardial infarctions in Sweden 1987–2010, https://www.socialstyrelsen.se/Lists/Artikelkatalog/Attachments/18511/2011-11-36.pdf (2011, accessed 28 April 2016).
[9] Mendis S, Thygesen K, Kuulasmaa K, et al. World Health Organization definition of myocardial infarction: 2008–09 revision. *Int J Epidemiol* 2011;40:139–46.
[10] Nissen SE, Tuzcu EM, Schoenhagen P, et al. Effect of intensive compared with moderate lipid-lowering therapy on progression of coronary atherosclerosis: a randomized controlled trial. *JAMA* 2004;291:1071–80.
[11] Brown TM, Deng LQ, Becker DJ, et al. Trends in mortality and recurrent coronary heart disease events after an acute myocardial infarction among Medicare beneficiaries, 2001–2009. *Am Heart J* 2015;170:249-80.

[12] Modig K, Andersson T, Drefahl S, et al. Age-specific trends in morbidity, mortality and case-fatality from cardiovascular disease, myocardial infarction and stroke in advanced age: evaluation in the Swedish population. *PLoS One* 2013;8:e64928.

[13] Swedish National Board of Health and Welfare. Myocardial infarctions in Sweden 1994–2014, https://www.socialstyrelsense/Lists/Artikelkatalog/Attachments/19943/2015-11-1pdf (2015, accessed 28 April 2016).

[14] Koch MB, Davidsen M, Andersen LV, et al. Increasing prevalence despite decreasing incidence of ischaemic heart disease and myocardial infarction. A national register based perspective in Denmark, 1980–2009. *Eur J Prev Cardiol* 2015;22:189–95.

[15] Manuel DG, Lim JJ, Tanuseputro P, et al. How many people have had a myocardial infarction? Prevalence estimated using historical hospital data. *BMC Public Health* 2007;7:174.

[16] Blais C and Rochette L. Trends in prevalence, incidence and mortality of diagnosed and silent coronary heart disease in Quebec. *Health Promot Chronic Dis Prev Can Res Policy Pract* 2015;35:184–95.

[17] Andersson T, Ahlbom A and Carlsson S. Diabetes prevalence in Sweden at Present and projections for year 2050. *PLoS One* 2015;10:e0143084.

[18] Andersson T, Ahlbom A, Magnusson C, et al. Prevalence and incidence of diabetes in Stockholm County 1990–2010. *PLoS One* 2014;9:e104033.

[19] Emilsson L, Lindahl B, Koster M, et al. Review of 103 Swedish healthcare quality registries. *J Internal Med* 2015;277:94–136.