Variability of cancer risk within an area: time to complement the incidence rate

Emanuele Crocetti, Francesco Giusti, Carmen Martos, Giorgia Randi, Tadeusz Dyba and Manola Bettio

The aim of this study was to show that age-adjusted cancer incidence rates for an area may not be representative of the incidence in subareas. We propose a simple measure to show the amount of geographical variability. European age-standardized incidence rates (ASRs) for 'all sites excluding nonmelanoma skin cancer', for men, in 2014, for Nordic countries as a whole, for each country (Denmark, Faroe Islands, Finland, Greenland, Iceland, Sweden and Norway) and for their regions, were retrieved from the Nordcan with corresponding standard errors SEs. We compared the ASR for Nordic countries versus single country and single country versus specific regions. The overlapping of 95% confidence intervals was used for ASRs comparisons. As a measure of variability, we computed the range between the highest and the lowest ASR within an area and the ratio between this range and the ASR of the overall area, $r/R = (\text{range}/\text{ASR}) \times 100$. The 95% confidence interval of the ASR for Nordic countries as a whole did not overlap those of the majority of the single countries; in fact, the r/R – which provides a clue for the amount of underlying geographical variability – was rather large (57.1%). Within countries, the variability was negligible in Iceland ($r/R = 9.6\%$), whereas the highest value was found in Sweden (37.1%). The ASR does not provide any information on underlying geographical variability. Therefore, its interpretation could be misleading. When data for subareas are available, the r/R, which is simple to compute and to understand, should be added to the ASR for providing more truthful information. *European Journal of Cancer Prevention* 26:442–446 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc.

Correspondence to Emanuele Crocetti, MD, European Commission, DG Joint Research Centre, Via E. Fermi, 2749 – TP127, 21027 Ispra (VA), Italy
Tel: + 39 0332 789926; fax: + 39 0332 783858;
e-mail: emanuelecrocetti@yahoo.com

Introduction

As a standard practice worldwide, population-based cancer registries (CRs) express the occurrence of cancer in a defined population in a certain period as the ratio between the newly diagnosed cancers and the at-risk resident population. This ratio is called the crude incidence rate (Boyle and Parkin, 1991). Cancer incidence increases with the ageing of the population. Therefore, incidence rates are strongly dependent on the age structure of the underneath population. Consequently, rates are computed using a standard age structure as a reference (age-standardized rate, ASR) to enable reliable comparisons across time and countries (Boyle and Parkin, 1991). Crude rates and ASRs are the standard indicators reported by all CRs independent of the size of the population at risk. These statistics are usually complemented by a measure of precision, the standard error (SE) of the rate and/or the 95% confidence intervals (CIs).

An incidence rate expresses the summary probability of developing cancer in the area covered by the CR. It provides no clues on the homogeneity or the heterogeneity of incidence rates across subareas.

To gain more insight into this topic and to explore the possible variability in ASRs among subareas of CRs, we analysed the data of the Association of the Nordic Cancer Registries, which makes data available in the Nordcan project (Engholm *et al.*, 2016).

Methods

We retrieved from the Nordcan the European ASRs for 'all sites excluding nonmelanoma skin cancer', for men, in 2014.

ASRs are available for three geographical layers as presented in Table 1: (a) Nordic countries as a whole; (b) single country: Denmark, Faroe Islands, Finland, Greenland, Iceland, Sweden, Norway; and (c) regions: five in Denmark: North Jutland, Central Jutland, Southern Denmark, The Capital and Zealand region; five in Finland: Helsinki, Kuopio, Oulu, Tampere and Turku region; two in Iceland: Reykjavik-Reykjanes and Outside The Capital;

Table 1 Data from Nordcan (*http://www-dep.iarc.fr/NORDCAN*): country layers, country/region name, number of incident cases of 'all sites excluding nonmelanoma skin cancer', in men, in 2014, resident population, European age-standardized incidence rates, SE of the ASR, lower and upper 95% confidence intervals and *r/R*

Layer	Area	*N*	Resident population	ASR (European)	SE	LCI	UCI	*r/R* (%)
1	Nordic countries	79 441	1 307 5123	453.1	1.6	449.9	456.3	57.1
2	Denmark	19 031	2 799 895	504.4	3.7	497.1	511.7	14.4
3	North Jutland	2004	292 697	474.8	10.9	453.4	496.2	
3	Central Jutland	4165	639 192	496.5	7.8	481.2	511.8	
3	Southern	4230	600 667	493.0	7.8	477.7	508.3	
3	The Capital	5245	860 818	507.1	7.1	493.2	521.0	
3	Zealand	3387	406 521	547.5	9.7	528.4	566.6	
2	Faroe Islands	78	25 039	251.0	28.8	194.6	307.4	
2	Finland	15 142	2 686 119	404.5	3.4	397.9	411.1	31.9
3	Helsinki	4849	922 582	427.2	6.2	415.0	439.4	
3	Kuopio	2194	403 920	345.5	7.7	330.5	360.5	
3	Oulu	1970	372 534	382.2	8.8	364.9	399.5	
3	Tampere	3811	545 749	474.6	7.9	459.1	490.1	
3	Turku	2303	441 350	348.5	7.5	333.7	363.3	
2	Greenland	99	29 742	384.1	40.6	304.5	463.7	
2	Iceland	694	164 257	387.2	14.8	358.1	416.3	9.6
3	Reykjavik	481	115 443	401.2	18.5	365.0	437.4	
3	Outside	213	48 818	364.1	25.4	314.3	413.9	
2	Norway	15 865	2 581 421	509.9	4.1	501.9	517.9	13.9
3	Central	2362	357 476	526.9	11.0	505.3	548.5	
3	Northern	1489	242 918	468.8	12.4	444.6	493.0	
3	South-Eastern	8685	1 433 445	502.9	5.5	492.2	513.6	
3	Western	3329	547 582	539.8	9.4	521.3	558.3	
2	Sweden	28 709	4 843 303	428.5	2.6	423.4	433.6	37.8
3	Northern	2521	444 391	364.5	7.6	349.6	379.4	
3	Stockholm-Gotland	6735	1 111 680	526.3	6.5	513.6	539.0	
3	Southern	5375	872 866	438.1	6.2	425.9	450.3	
3	South-Eastern	3130	510 943	415.8	7.7	400.6	431.0	
3	Uppsala-Örebro	5852	1 002 193	385.8	5.3	375.5	396.1	
3	Western	5096	901 258	407.1	5.9	395.6	418.6	

The smaller the *r/R* the lower the variability across subareas ASRs.
range/Rate% (*r/R*): range is the absolute difference in ASRs between the greatest and the lowest ASR of subareas in the lower layer; rate is the ASR.
ASR, age-standardized incidence rate; LCI, lower 95% confidence interval; UCI, upper 95% confidence interval.

six in Sweden: Northern, Stockholm–Gotland, Southern, South-Eastern, Uppsala-Örebro and Western region; and four in Norway: Central, Northern, South-Eastern and Western region.

The overall male resident population in Nordic countries in 2014 was 13 075 123. Residents in single countries and regions are shown in Table 1.

ASRs express the number of new cases diagnosed among 100 000 men in 2014 according to the observed age-specific rates and the age-groups of the European standard population.

We also retrieved from Nordcan the SE of the ASRs and we computed the 95% CIs according to the method of the binomial approximation (Boyle and Parkin, 1991) (Table 1).

We evaluated whether two rates were different inspecting the overlap between specific 95% CI (Schenker and Gentleman, 2001). The precision of the age-specific rates that concur in the calculation of ASR increases when the number of cases in this group increases. This applies to each age group and thus to ASR as the whole entity. The overall numbers observed yearly in the analysed series (Table 1) were, with the exception of Faroe Islands and Greenland, in the order of several hundreds or even

thousands. SEs of the ASR are greater when the numbers on which they are based are small.

We compared the ASR at each geographical level with the level underneath: Nordic countries versus Denmark, Faroe Islands, Finland, Greenland, Iceland, Sweden, Norway and single country versus specific regions.

Moreover, we computed the absolute difference (range) between the highest and the lowest ASR within a nested layer (the range between countries for Nordic countries and between regions for a specific country). Then we calculated the percent ratio between this range and the ASR of the level above interpreted as a summary value of the subareas (for Nordic countries or a single country, respectively), $r/R = (\text{range}/\text{ASR}) \times 100$.

The *r/R* provides a measure of the variability across the available ASRs of the nested level for which the ASR represents the summary measure. The smaller the *r/R* (minimum 0%), the lower the variability across subarea ASRs.

Results
In Fig. 1, the ASRs for Nordic countries as a whole, for single countries and for country-specific regions, are shown with the corresponding 95% CI. The ASRs for countries appear to be scattered in the picture. In fact,

Fig. 1
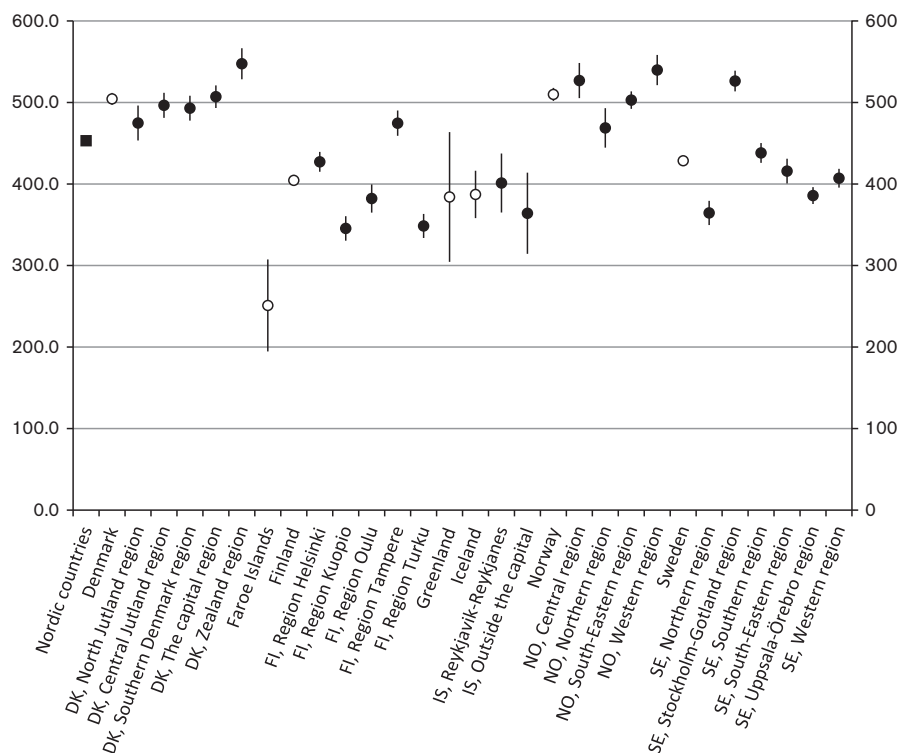
European age-adjusted incidence rates for 'all sites excluding nonmelanoma skin cancer', for men, in 2014, for Nordic countries, single countries and regions. From Nordcan (*http://www-dep.iarc.fr/NORDCAN*).

the ASR for Nordic countries (453.1) is compatible with the Greenland's one only (384.1) because of the wide range of variability of the latter, because of the small number of cases on which it is based and the resulting imprecision in its computation (wide 95% CI). In contrast, Denmark (504.4) and Norway (509.9) showed greater ASRs that the Nordic countries one and Faroe Islands (251.0), Finland (404.5), Iceland (387.2) and Sweden (428.5) have lower values than the supranational summary ASR (Fig. 1).

In Table 1, for each area (Nordic countries, country and region), the number of cancer cases for 'all sites excluding nonmelanoma skin cancer', for men, in 2014 and the resident population are reported together with the ASR, the SE and the 95% CI.

The inverse relationship between number of observed cases and the SE is evident. In fact, SE is only 1.6 (cases per 100 000 men in 2014) for Nordic countries (on the basis of 79 441 analysed cases), whereas it is 40.6 for Greenland (99 cases).

In Table 1, the r/R is also reported for geographical level 1 (Nordic countries vs. countries) and 2 (single countries vs. regions).

When the ASR of Nordic countries is evaluated together with the r/R, the value of $r/R = 57.1\%$ provides a clear hint of a huge intercountries variability in ASRs, clearly shown in Fig. 1. In fact, this r/R means that the range between the lowest and the highest country-specific ASR is almost 60% of the Nordic country ASR.

Also within single countries, the overall ASR may not represent the regional ASRs and the amount of internal variability (Fig. 1) is well described by r/R (Table 1).

The smallest r/R value (9.6%) was observed in Iceland, where the small numbers of observed cases led to a non-negligible uncertainty in the regional estimates whose wide 95% CI overlapped the national one. A minor amount of variability ($r/R = 13.9\%$) was present in Norway, where the Northern region (ASR = 468.8) had a lower value and the Western region (539.8) had a higher ASR than the summary one. Almost the same r/R was present in Denmark (14.4%), where North Jutland (474.8) showed an ASR lower than the national value and Zealand (547.5) showed a greater one. Finland showed a greater inter-regional variability (31.9%), with Kuopio (345.5) and Turku (348.5) below and Helsinki (427.2) and Tampere (474.6) above the national mean. Finally, the slightly higher internal variability was found in Sweden (RR = 37.8%) where three regions, Northern

(364.5), Uppsala-Örebo (385.8) and Western (407.1), were below the national ASR and Stockholm–Gotland (526.3) higher than the country one.

Conclusion

This epidemiological exercise underlines that ASRs, which clearly provide the level of cancer incidence in a specific area and time for geographical and time comparisons, do not provide any information on possible internal variability. In fact, the SE, which usually accompanies ASR, refers only to the precision of the estimate and does not reflect the possible heterogeneity in cancer incidence in the area.

Therefore, the ASR of a CR, although correct from the computational point of view, and informative for geographical and time comparisons, could represent the incidence level only in some subareas or even in none.

If a CR also provides ASR for subareas, r/R is not necessary because the information on possible geographical heterogeneity is available. In contrast, if a CR only publishes a summary ASR, as happens for many CRs in Cancer incidence in five continents (Ferlay *et al.*, 2014), which is the most well-known and authoritative publication in the field, r/R is invaluable to have a clear impression of the variability behind the ASR.

When incidence data are available for different geographical layers, it is possible to add to the ASR a summary measure about the underlying variability. The Nordic countries dataset provided the invaluable chance of evaluating three subgeographical levels: supranational, national and regional.

We propose to compute the range between the highest and the lowest underlying ASRs to divide it by the ASR (r/R) and to express the result as a percentage.

The index r/R has been chosen among other more formal statistics (e.g. extreme quotient) (Gumbel and Keeney, 1950) because it only relies on ASRs and provides a direct measure of the effect of internal heterogeneity (range between maximum and minimum ASR in subareas) on the overall summary ASR.

In our example, on the basis of long-standing high-quality Nordic countries incidence data (Ferlay *et al.*, 2014), the r/R for the Nordic countries was quite high (57.1%), suggesting that the national ASRs could vary notably. In fact, the overall ASR for Nordic countries did not correspond with any of the national ASRs, out of Greenland's one (Fig. 1).

Also at a national level, when regional estimates are available, it is possible to add to the national ASR the r/R based on regional ASRs to express how well the national ASR represents the regional ones. In the dataset analysed, we showed that country ASR may reflect more (Iceland, Denmark and Norway) or less accurately (Finland and Sweden) the incidence of cancer in the different regions within a country.

The comparison between ASRs using the 95% CI overlap is simple and intuitive (Schenker and Gentleman, 2001) and showed major differences in ASRs between and within areas.

This study was based only on one incidence year. To check the reliability of r/R, we repeated the exercise also for the year 2012. The r/R in 2012 were similar to that in 2014 (data not shown) for almost all the countries, with the exception of Sweden, for which r/R showed in 2012 a smaller heterogeneity ($r/R = 16.6\%$) than in 2014 (37.8%). The reason for this strong change was the change in the incidence ASR in the Stockholm–Gotland region from 2012 (420.5 cases/100 000) to 2014 (526.3). This change was the effect of a study on prostate cancer carried out in the county between 2012 and 2014 (Grönberg *et al.*, 2015). The ASR for all causes except skin and prostate cancer were 268.3 and 266.7, respectively. This example confirms that r/R reflects the true variability within an area.

Heterogeneity was identified among countries (areas between around 25 000 and 4 800 000 resident men) and among regions of several hundred thousand inhabitants, except for Iceland, where the population is smaller than in any of the other countries with regional information available.

It is possible to identify slight differences in cancer incidence between two geographical areas if the number of cases (population) is huge. Then, the ASRs are precise and the 95% CI is narrow. Thus, it is easier to detect a slight difference between two large (populated) regions than between two small ones. For example, between the ASRs of Kuopio and Oulu (highly populated), there is the same difference as that between the two Icelandic regions (poorly populated), but only the first two do not have overlapping CIs.

In general, the unavailability of a unique population-unit for subareas (countries, regions, provinces, counties, etc.) makes comparisons across areas difficult.

With the increase in the number of subareas, the variability among them is expected to increase and consequently the r/R. The aim of r/R is exactly to offer summary and straightforward information on possible outliers. In case r/R is small ($\sim <10$–15%) it is immediately clear that all the ASR for each of the subareas are concentrated in a quite narrow range and if it is large ($>30\%$) it underlines that at least one of them is rather different from the overall ASR.

The r/R is a measure intended as a macro indicator of major heterogeneity among quite large subareas (e.g. regions in a country). For small areas and cluster analysis, other methods have to be chosen (Colonna and Sauleau, 2013).

CRs should start to provide also general information on internal cancer incidence geographical variability in

addition to standardised incidence rates. This would make the information more complete and clear for readers, avoiding misinterpretations. When incidence for subareas is available, r/R, which is very simple to compute, could be presented together with the general ASR as a first attempt to raise the issue.

The interpretation of incidence ASR requires the combined reading of ASR, SE and r/R: the ASR shows the level of incidence, the SE shows the precision of the ASR and r/R shows the amount of internal geographic variability. The r/R will be smaller if the ASR for subareas are quite similar to each other (more or less precisely estimated) or greater if they are rather different. This is the original and useful contribution provided by the r/R.

Acknowledgements
Conflicts of interest

References

Boyle P, Parkin DM (1991). Statistical methods for registries. . In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet, RG, editors. *Cancer registration: principles and principles* . Lyon, France: IARC Scientific Publication (No. 95). pp. 126–157.

Colonna M, Sauleau EA (2013). How to interpret and choose a Bayesian spatial model and a Poisson regression model in the context of describing small area cancer risks variations. *Rev Epidemiol Sante Publique* **61**:559–567.

Engholm G, Ferlay J, Christensen N, Kejs AMT, Hertzum-Larsen R, Johannesen TB, , et al. 2016 NORDCAN: cancer incidence, mortality, prevalence and survival in the nordic countries, version 7.3 (08.07.2016). Association of the Nordic Cancer Registries. Danish Cancer Society. Available at: *http://www.ancr.nu*. [Accessed 19 December 2016].

Ferlay J, Bray F, Steliarova-Foucher E, Forman D (2014). *Cancer incidence in five continents, CI5plus IARC CancerBase No 9*. Lyon, France: International Agency for Research on Cancer.

Grönberg H, Adolfsson J, Aly M, Nordström T, Wiklund P, Brandberg Y, *et al.* (2015). Prostate cancer screening in men aged 50–69 years (STHLM3): a prospective population-based diagnostic study. *Lancet Oncol* **16**: 1667–1676.

Gumbel EJ, Keeney RD (1950). The extremal quotient. *Ann Math Stat* **21**: 523–538.

Schenker N, Gentleman JF (2001). On judging the significance of differences by examining the overlap between confidence intervals. *Am Stat* **55**:182–186.