The denominator in general practice, a new approach from the Intego database

Stefaan Bartholomeeusena, Chang-Yeon Kimb, Raf Mertensc, Christel Faesb and Frank Buntinxa


“A practice population is a nebulous concept: it is undefinable and cannot consistently be estimated.” SJ Kilpatrick

Background. To determine the denominator or the ‘population at risk’ is a problem which has long been encountered in general practice-based epidemiological research. It is important for calculating epidemiological figures.

Objectives. The aim of this article is to demonstrate how in the absence of a patient list, a reliable denominator can be calculated, starting from the number of patients who contacted their GP in the period of one year. Therefore a brief overview will be given from known approaches, then the new approach will be illustrated on a database named Intego, with data from 43 general practices in Belgium.

Methods. The Intego database contains information about patient contacts, diagnoses, laboratory results and drug prescriptions, extracted from the participants’ structured electronic medical record system. The number of patients who contacted the practice in a year can be calculated from the Intego data. On the other hand, the percentage of the population that consults a GP during a particular period was obtained from the reimbursement claims data available from the sickness funds. By combining these two datasets, stratified by age, gender and district, a correction factor was calculated. An estimate of the real size of the Intego practice populations was obtained by extrapolating the yearly contact group by this factor.

Results. In 2003 according the Intego-register, 64 161 patients contacted their family practice and this correlated with an estimated practice population of 80 094 patients. The absence of the socio-economic status in the estimation is irrelevant in our model of estimating the practice population.

Conclusion. The availability of a denominator in general practice-based research is essential to calculate epidemiological figures. This method using a correction factor makes it possible to calculate a reliable practice population. A similar approach will probably also be applicable in other European countries.

Keywords. Data collection, epidemiology, family practice, medical records, population.

Introduction

Various rates can be used to describe the occurrence of a disease in a population. In most cases, the number of events (the numerator) must be divided by the size of the population, which is the denominator or ‘population at risk’. The number of events in the numerator is generally relatively simple to determine, but in the absence of a population to be related to, it is of limited value. Moreover, this figure does not allow comparisons with other populations.

It is usually easy to obtain the number of different patients who consulted their GP in a given period; this is the yearly contact group (=YCG). But in countries without capitation or other systems linking patients to a given GP it is much more difficult to know how many other people (the so-called zero class) are also to be counted as belonging to a given practice population (PP).
This problem is known in the literature as the denominator problem. Since people are not registered with a particular GP in Belgium and may consult several different doctors, the doctors do not have patient lists or practice populations. Consequently, research projects almost invariably are confronted with this denominator problem.

In the past there have been a number of attempts to calculate this denominator. In 1994, a European project tried to identify an optimum denominator, to be used in various countries. After having evaluated all available methods, FW Schwartz concluded that the YCG was the only realistic denominator, although not a perfect one.

In this paper, we will examine which of the available methods is useful for defining a practice population in our setting. Next, we will present a new approach for obtaining a reliable estimate of the population at risk on the basis of the YCG.

**Literature review**

Our literature search included both indexed and non-indexed journals. Additional to a PubMed search, we therefore hand searched the content pages of Allgemeinmedizin International and used formal and informal databases of Huisarts en Wetenschap and Huisarts Nu. Literature was from references cited in the articles and research reports. We also used information of a number of grey literature reports, including the European Community study report of Schwartz.

Four statistical models have received the most attention in the literature. Using the data from the national morbidity study in the UK for 1971–1972, SJ Kilpatrick looked at the distribution of episodes of illness. An episode of illness is defined here as a period of illness in which one or more contacts with the GP are required. He determined the relative frequency of the number of patients with 1, 2, 3 etc. episodes of illness per year and concluded that they followed a geometric distribution. From this it is possible to calculate the group of patients with zero episodes. Extensive research using this method in the USA, however, yielded disappointing results. One also should question whether an extrapolation to the zero class, as it is outside the range of observations, is appropriate.

P Krogh-Jensen started out from the number of consultations by a patient per year in 216 practices in Denmark in 1977 and 1978. He described the distribution as a double negative binomial distribution, in which patients were subdivided into two groups. The first group consisted of patients who only sporadically consulted the practice and the second group consisted of patients who consulted the practice regularly. Despite the fact that a huge amount of data is involved, still some comments remained. The group of children under 16 was not included in the study. The calculation itself is quite complex and not very clear.

On the basis of the same Danish data E Schach studied whether the Poisson log normal distribution could be a model for determining the practice population. She concluded that this distribution could be a model for defining the practice population. In this analysis children aged under 16 were once again not included and the distribution of expected against observed non-consulters only corresponded at frequencies of between 30 and 40%.

BW Smith worked out a model which he later improved and in which he started from the number of contacts per year per patient. The result is an odds ratio of consulting the practice more than a specified number of times per year. The model was reported to have a deviation of less than 2% when calculating the practice population. Interpretation of the odds ratio is difficult, however, and there are also methodological criticisms of this model.

As early as 1976 J Garson proposed a ‘correction factor’ to convert the YCG to the PP, based on the observation that a fixed proportion of the population consult their GP during a given period. This information had been provided by the medical insurance institution with which the patients were registered. In 1982, D Cherkin further elaborated the method by proposing a specific correction factor per age group and gender.

In the Intego database, the yearly number of contacts between a patient and their GP is not known, so the models used by P Krogh-Jensen, E Schach and BW Smith cannot be used. In the method used by SJ Kilpatrick the validity of an extrapolation beyond the area of observation can be questioned or whether the behaviour of non-consulters can be equated with that of those who do consult. With this method there is also a problem with patients who consult without a new diagnosis being made.

**Methods**

*The Intego family practice registration network*

Intego is the name of a registration network for family practices in Flanders. Anonymised diagnoses, laboratory results and drug prescriptions are systematically recorded in a central database. In 2003, a total of 43 practices (51 GPs) located throughout Flanders collaborated in the data collection process. The quality of the data received from the practices is checked by different parameters. The number of new diagnoses per patient must be greater than one, the size of the practice must be stable, diagnoses have to be registered by coded keywords. The GPs receive feedback and instructions how to register diagnoses properly.

From this Intego database, the number of cases of a particular disease and the YCG can easily be obtained, but not the size of the entire population at risk. This YCG is determined and stratified by age, gender and residence.
The second source of information was the reimbursement claims database of the national compulsory health insurance. The latter is administered by seven not-for-profit sickness funds, who have recently pooled some of their data in the Intermutualistic Data Agency (IMA). The sickness funds have complete information on all reimbursed medical and paramedical acts, including GP consultations. Their total population of insures is of course also perfectly known, including age, gender, residence and eligibility under a priority scheme. This is a system whereby patients who are retired, widows or widowers, orphans or disabled and on low incomes receive an increased reimbursement of their medical costs. It can be seen as an approximate indicator of socio-economic status. From this data we can determine the proportion of the population who consulted their GP in the period of one year. This proportion can be stratified by age, gender, residence and socio-economic status, which is the correction factor. By dividing the YCG by the correction factor we get the PP.

To estimate the effect of socio-economic status, we modelled the YCG/PP ratio in the IMA data using a negative binomial distribution. From a previous study\(^1\) this was shown to be the most appropriate for our objective. In a first model age, sex and district were used as covariates together with all possible interaction terms. In a second model we added socio-economic status together with the additional interaction terms providing the best fit, based on a stepwise selection procedure.

**Results**

*Information from Belgian databases*

In the Belgian National Health Interview Survey,\(^1\) 10 000 people were questioned about their use of health care during the year 2001. 81% of the Belgian population declared to have had at least one contact with their GP during the year. (77% in men and 84% in women).

According to the IMA figures for 2001, 77% of the population consulted a GP at least once during the year: 81% of women and 73% of men. These figures, however, differ according to age, gender and socio-economic status. A comparison of the figures from the National Health Interview Survey and the IMA shows that they correspond quite closely (Fig. 1). From this information it can be concluded that the YCG is approximately 80% of the PP.

In Figure 2 the percentage of the population from the IMA database having consulted their GP in 2001 is shown in the 22 Flemish districts. The urban areas of Halle-Vilvoorde, Gent and Antwerp have the lowest percentage of patients consulting their GPs in 2001, as could be expected from the higher concentration of hospitals and consultants. Overall, the percentage fluctuates around 80%.

In Figure 3 the percentage of consulting patients in the Flemish population is analysed by age group and gender. Up to the age of 14 years, boys and girls consult the same proportions, but in the young adult category women consult significantly more often than men. This difference declines from the age of 60 onwards and very elderly people almost all consult their GP at least once annually. In older age groups the YCG therefore approaches the PP.

*Calculating the PP from the YCG in the Intego database*

On the basis of this stratified data from IMA, it is possible to calculate a PP using the data from the Intego database. In this database it is not known whether or not the patient is covered by the priority scheme system and this factor is therefore not included in the calculation.
The YCG is determined for each district, 5 year age group and gender. This stratified YCG is then divided by the percentage of consulting patients in the relevant stratum from the IMA data, which gives the PP for that cell. In Table 1 an example is given of the calculation using the figures for 2001. By adding up all cells, the total PP for a given year can then be obtained, taking into account three of the four factors that have a significant influence on its size, namely age, gender and region, but not socio-economic situation.

It therefore is necessary to estimate the error that is caused by the absence of socio-economic status in the stratification. From the IMA data we know that socio-economic status is an important predictor of the YCG/PP ratio. The odds ratio of belonging to the YCG in patients of the high versus low socio-economic status group is 0.84. As socio-economic status is highly related to age, sex and district, being the three available characteristics, it is not the crude or even the adjusted impact of socio-economic status that is of interest to us.
As SJ Kilpatrick says in the introductory quotation, the PP is not a clearly circumscribed entity. A real denominator only exists in those countries where patients have a fixed registration with a particular GP, and so the practice is defined. What is more, these patient lists are constantly changing due to arrivals and departures, and there is often a significant time delay between the date when a change occurs and the date when it is entered in the register. Countries with a registration system of this kind are, however, the exception rather than the rule. What is more, the statistical models found in these countries cannot simply be applied to other countries with a different type of health care organisation. Defining a practice population therefore means calculating a quantity without any possibility of asserting that it is correct. It is, however, possible to define factors that can be assumed to influence whether or not people contact their GP within a particular period.

None of the mathematical models are currently used for calculating incidences. Moreover, there are innumerable reasons why a patient does or does not consult a doctor for a particular complaint. The way the patient experiences the symptoms, the socio-cultural environment in which he lives and the organisation and accessibility of health care are all factors which are significant in the decision to consult or not consult. It is doubtful whether such a complex set of factors can be summarised in a mathematical formula.

In the final report of the European project set up to define a denominator, FW Schwartz states that the YCG is not the perfect denominator, but that it is the only realistic one. Nevertheless, it cannot be used for international comparisons.

The Intego register
In 2003, the Intego register had a YCG of 64,161 patients (33,383 women and 30,778 men). The corresponding estimated PP amounts to 80,094, made up of 40,399 women and 39,695 men. The age and gender distribution of this PP are close to the figures for the Flemish population.
result of computerisation, defining the YCG for a practice has become a simple task. On the basis of the IMA figures, which can be seen as the standard at the group level for the Belgian situation, the age group, gender and district are taken into year to year. Moreover, as the marginal effect of socio-economic status on top of the available information was found irrelevant for our objectives, the method could be adopted for routine use with the Intego registry. The method is likely to be also suitable for general practice-based morbidity registries in other European countries without a list system, provided complete insurance data or comparable information is available. Further study should show whether socio-economic status would not become a relevant co-variable when comparing different subgroups (e.g. regions) within our population.

Certain precautions and assumptions have to be made in case the method of the correction factor is used. To avoid that the same patients are counted twice (or more) in nearby practices, the distance between practices has to be large enough. The change in the YCG of a practice may not be too abrupt from year to year. So it is not possible to use the correction factor for practices which are rapidly losing or gaining patients.

Another assumption is that almost all people living in a region must be insured by a health insurance or that insured and non-insured people tend to contact their GP in a similar way. Unless for illegal inhabitants, this is largely the case in Belgium, but may not be so in all European countries.

We developed a method that seems to provide a good estimate of the practice population in a country where no practice lists are available. This may be an important step for general practice-based epidemiology in Europe.

Acknowledgements

Thanks go to all collaborating GPs and to the IMA staff members. Without their co-operation this work would not be possible. Thanks go also to Prof Dr Jan Vandebroucke from the Department of Clinical Epidemiology, Rijksuniversiteit Leiden, The Netherlands for his advice and critical discussion of a previous version of this paper.

Declaration

Funding: the Intego project is being carried out with support from the Government of Flanders and was commissioned by its Health Minister.

Ethical approval: the project was approved by the Medical Ethical Committee of the Faculty of Medicine of the K.U.Leuven.

Conflicts of interest: none.

References

3 Schwartz FW. The European denominator project. Comparison and Harmonisation of Denominator Data for Primary Health Care Research in Countries of the European Community. End Report. Hannover Medical School, Department of Epidemiology and Social Medicine; 1994, 140.
7 Schach E. Comparison between practices; Estimating the size of the practice population. Allgemeinmedizin International 1983; 8: 141–144.