Postprint Version	1.0
Journal website	http://www.schattauer.de/index.php?id=1268&pii=me08020098&no_cache=1
Pubmed link	http://www.ncbi.nlm.nih.gov/pubmed/18338080?ordinalpos=1&itool=EntrezSystem2.PEntrez.Pubmed.Pubmed_ResultsPanel.Pubmed_RVDocSum
DOI	10.3414/ME0471

This is a NIVEL certified Post Print, more info at http://www.nivel.eu

Estimating Morbidity Rates from Electronic Medical Records in General Practice: Evaluation of a Grouping System.

M. C. J. Biermans[1], R. A. Verheij[2], D. H. de Bakker[2], G. A. Zielhuis[3], P. F. de Vries Robbé[1]

1Department of Medical Informatics, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands
2Netherlands Institute for Health Services Research, Utrecht, the Netherlands
3Department of Epidemiology and Biostatistics, Radboud University Nijmegen Medical Centre, Nijmegen, the Netherlands

Summary

Objectives: In this study, we evaluated the internal validity of EPICON, an application for grouping ICPC coded diagnoses from electronic medical records into episodes of care. These episodes are used to estimate morbidity rates in general practice.

Methods: Morbidity rates based on EPICON were compared to a gold standard; i.e. the rates from the second Dutch National Survey of General Practice. We calculated the deviation from the gold standard for 677 prevalence and 681 incidence rates, based on the full dataset. Additionally, we examined the effect of casebased reasoning within EPICON using a comparison to a simple, not case-based method (EPI-0). Finally, we used a split sample procedure to evaluate the performance of EPICON.

Results: Morbidity rates that are based on EPICON deviate only slightly from the gold standard and show no systematic bias. The effect of case-based reasoning within EPICON is evident. The addition of case-based reasoning to the grouping system reduced both systematic and random error. Although the morbidity rates that are based on the split sample procedure show no systematic bias, they do deviate more from the gold standard than morbidity rates for the full dataset.

Conclusions: Results from this study indicate that the internal validity of EPICON is adequate. Assuming that the standard is gold, EPICON provides valid outcomes for this study population. EPICON seems useful for registries in general practice for the purpose of estimating morbidity rates.

INTRODUCTION

Prevalence and incidence rates in general practice are used by scientists to monitor symptoms and diseases in the population, by policymakers to formulate and to evaluate health care policy, and by general practitioners to make probability diagnoses. Under certain conditions, listed below, the data from electronic medical records (EMRs) in general practice are a valuable source for estimating these rates.

A necessary condition for estimating morbidity rates from EMRs is that patient lists are available. These lists are needed to determine the size of the underlying practice population, i.e., the denominator of the morbidity rates. Countries with a list system, such as the Netherlands, Denmark, the United Kingdom, Italy, Spain, Portugal, and Slovenia, are also countries with a gatekeeping system [1]. The presence of such a system is a very favorable condition, because within this system, the main pathway to medical care is through general practice. Hence, morbidity rates in general practice within these systems provide a good indication of the health status of the general population [2, 3].

Another requirement is that general practitioners record coded diagnoses in their EMRs. To estimate incidence rates, knowing whether diagnoses refer to a new or a continuing health problem is an additional prerequisite. Furthermore, all diagnoses need to be grouped into episodes to estimate the numerator of the morbidity rates. An episode of care includes "all encounters for the management of a specific health problem"[4]. Consider, for example, a patient who consults the general practitioner for a tension headache (diagnosis a), which, the next day, develops into a migraine (diagnosis b). Most likely, both diagnoses refer to the same health problem which, when estimating morbidity rates, should be counted only once, namely as a case of migraine. To avoid double counting, the *two* diagnoses need to be grouped in *one* episode of care named 'migraine'.

Diagnoses can be grouped into episodes, either directly by the general practitioner, or afterwards, through manual review or use of a computerized method. In an earlier study [5], we described the development of an application called EPICON, which can be used for the grouping of coded diagnoses from EMRs in general practice into episodes of care. EPICON renders it possible to estimate morbidity rates from EMRs in general practice without having to resort to expensive and time-consuming manual review.

EPICON is based on a combination of logical expressions, a decision table, and information extracted from manually grouped diagnoses by case-based reasoning. Previous research on the process of grouping diagnoses shows that EPICON groups 95% of all diagnoses correctly [5]. The present study focuses on how the grouped diagnoses translate into morbidity rates. In this study, we will examine if EPICON serves its purpose by comparing morbidity rates based on EPICON to a gold standard.

The prevalence and incidence rates from the second Dutch National Survey of General Practice (DNSGP-2) are considered as the gold standard [6-8]. In the DNSGP-2, diagnoses from 89 general practices were grouped afterwards into episodes by a semicomputerized method. This method consists of a computerized component, in which diagnoses are grouped automatically, and a manual component, in which diagnoses are grouped by medical coders.

The development of EPICON started with a simple, fully computerized method, called EPI-0. This variant is identical to the computerized component of the semi-computerized method, but instead of referring 'difficult diagnoses' to the manual component, it groups all remaining diagnoses automatically as separate episodes. EPICON groups these remaining diagnoses by case-based reasoning. In this process, EPICON imitates the manual grouping by retrieving and reusing information and knowledge from the DNSGP-2 dataset. For instance, EPICON solves the grouping problem in the example above by counting how often, in the DNSGP-2 dataset, manually grouped combinations of migraine and tension headache were placed together into one episode. The appendix provides a brief description of EPICON and EPI-0.

The main aim of this study is to assess the internal validity of EPICON. Internal validation refers to the performance of a system in a sample used to develop the system. Internal validation is a requirement for external validation, which refers to the performance of a system in a new sample of patients [9].

In this study, we will address the following research questions: 1) What is the deviation from the gold standard for morbidity rates based on EPICON? In particular, what is the effect of case-based reasoning on the deviation from the gold standard? 2) What is the performance of EPICON among patients not included in the development of EPICON but from the same population?

METHODS

Dataset

We used data from EMRs of practices associated with the Netherlands Information Network of General Practice (LINH) [10, 11]. The general practitioners within this network assign diagnosis codes to consultations, prescriptions, and referrals. The diagnoses are coded according to the International Classification of Primary Care (ICPC, first edition) [12]. This classification system includes 685 different diagnosis codes, classified into 17 chapters each of which refers to a specific body system or problem area. Within each chapter, codes 1-29 refer to symptoms and complaints, and codes 70-99 refer to diseases.

Each consultation diagnosis is characterized as either belonging to a new or an ongoing 'type of episode'. A new episode refers either to a newly presented health problem or to a recurrent health problem, while an ongoing episode describes a continuing health problem.

Representativeness

The 89 practices are representative of all Dutch general practices with respect to region of residence (i.e. the province in which the practice is located), dispensing status (i.e. whether the practice is permitted to dispense its own prescriptions), and degree of urbanization. Solo practices, however, are slightly underrepresented in this sample (53% versus 64%) [13]. The patient population comprises a representative sample of 2% of the Dutch population regarding age, gender, and type of medical insurance [14].

Episode Construction

Within the framework of the DNSGP-2, episodes were constructed afterwards for 89 LINH practices for a period of one year (2001). Episodes were named after the last disease code within an episode. If no disease code existed, the episode was named after the last symptom code within an episode. In this study, we regrouped all diagnoses in these 89 practices into episodes using both EPI-0 and EPICON. Consequently, we had one dataset with three different episode constructions: 1) the existing DNSGP-2 episodes (the gold standard), 2) the EPI-0 episodes, and 3) the EPICON episodes [5].

Split-sample

We used a split-sample approach to examine the performance of EPICON.This approach allows for overestimation, a well-known statistical problem where a model is fitted and evaluated using the same dataset [15]. We randomly split the group of patients with diagnoses (n = 256,227) in half.We used half of the dataset (the training set, n = 128,113 patients) to construct the case bases used by EPICON to group diagnoses into episodes. We then used the other half of the dataset (the test set, n = 128,114 patients) to test EPICON, i.e., the diagnoses of these patients were grouped into episodes based on the split-sample case bases. This split-sample procedure adds a fourth episode construction.

The obvious disadvantage of the splitsample technique is its inefficiency, i.e., the training set only makes up half of the total dataset. To quantify this inefficiency, we made a fifth episode construction, in which we applied the split-sample case bases (based on the training set) to the training set. In this way, we could 'separate' the intertwining effects of

overestimation and inefficiency in the test set. We would know that overestimation (and not inefficiency of the splitsample approach) was a problem if the training set would show less deviation than the test set.

Prevalence and Incidence Rates

The prevalence rate is defined as the proportion of the population with a disease during the one-year study period. The incidence rate refers to the occurrence of new episodes of disease during the observed person-years at risk.

We used the episode names that resulted from the five episode constructions for the numerator of the morbidity rates. Eight practices did not record exactly 365 days, because of vacation, sick leave, etc. For that reason, the episodes were weighted for the number of days recorded (i.e. 365/number of days recorded). For the numerator of the prevalence rates, we counted, per episode name, the number of patients with at least one new or ongoing episode. For the numerator of the incidence rates, we counted, per episode name, the number of new episodes.

The mid-year population (i.e., the average of the population at the beginning and the end of the one-year study period, n = 343,853) of the 89 general practices was used as the denominator. Half of the midyear population was used to calculate the split-sample morbidity rates. Morbidity rates of symptoms and diseases that occur only in the female (W, X) or the male chapters (Y) are based only on the female or the male mid-year population.

A total of 677 different prevalence rates and 681 different incidence rates were calculated for each of the five episode constructions.

Some of the 685 possible rates were excluded, because they do not refer to a symptom or disease (‘no disease’ (A97)), they refer to incident events only (‘perinatal mortality’ (A95), ‘death’ (A96), ‘(un)complicated labor/delivery of live/stillbirth (W90-W93)), or no cases (‘trachoma’ (F86)) or no new cases occurred (‘neoplasm cardiovascular’ (K72), ‘syphilis female’ (X70)).

[FIGURE 1]

Comparison of Morbidity Rates

Figure 1 shows the design of the study. We calculated the differences between the rates based on EPI-0, EPICON, the test set or the training set on the one hand, and the gold standard on the other hand. We compared the deviation of rates based on EPI-0 to the deviation of rates based on EPICON to examine the effect of case-based reasoning. The comparison between the deviation in the test set to the deviation in the training set provides insight into the separate effects of overestimation and inefficiency.

We distinguished between infrequently (rate < 1 per 1000 per year) and frequently (rate ≥1 per 1000 per year) occurring symptoms and diseases (rates are based on the gold standard). This divided the large number of rates into two approximately equal parts. For infrequently occurring symptoms and diseases, we calculated the absolute differences between the morbidity rates (absolute deviation). For symptoms and diseases that occurred frequently, we calculated the percentage to which the rates differed from the gold standard (relative deviation).

RESULTS

Morbidity Rates Based on EPICON

Figures 2-5 show the morbidity rates based on EPICON in comparison to the gold standard. In general, these morbidity rates deviate slightly from the gold standard and show no systematic bias.

Infrequent prevalence rates (Fig. 2) deviate from – 0.04 (‘hypoglycemia’ (T87)) to 0.06 (‘fear of other respiratory disease’ (R27)). An outlier is the diagnosis ‘investigation with

abnormal results' (A91)). The prevalence of this diagnosis is 0.94 per 1000 patients based on EPICON (read from the x-axis of Fig. 2) and 0.85 per 1000 patients according to the gold standard, which is a difference of 0.09 (read from the y-axis of Fig. 2). EPICON probably failed to group this diagnosis correctly in a number of cases, because it is a very non-specific diagnosis that can be linked to many different diagnoses. Frequent prevalence rates (Fig. 3) deviate from – 6% ('vomiting/nausea of pregnancy' (W05)) to 12% ('abnormal unexplained blood test' (B85)). The deviation for infrequent incidence rates (Fig. 4) ranges between – 0.05 ('cervicitis/ other cervical disease' (X85)) and 0.07 ('hypertension with involvement of target organs' (K87)). Outliers are 'generalized/ unspecified pain' (A01), which is difficult to group because of its non-specificity, and 'osteoporosis' (L95). EPICON grouped the latter incorrectly in a number of cases because 'osteoporosis' (L95) can be linked to many other diagnoses in the musculoskeletal chapter (L), and occurs mainly in elderly patients who have a great deal of comorbidity. The deviation for frequent incidence rates (Fig. 5) varies from – 5% ('acute myocard infarction' (K75)) to 9% ('asthma' (R96)). An exception is 'emphysema/COPD' (R95)). The incidence rate of this diagnosis based on EPICON is 17% higher than the gold standard (1.78 versus 1.52 per 1000 person- years). The reason for this difference could be that 'emphysema/COPD' (R95) occurs frequently in elderly patients with concurrent diseases such as 'decompensatio cordis' (K77), 'pneumonia' (R81), 'chronic bronchitis/bronchiectasis' (R91), and 'asthma' (R96). The intricate distinction between exacerbations (usually grouped within an ongoing episode) and complications (usually grouped as a separate episode) within this group of diseases may have caused variation in the manual (the gold standard) and subsequently also in the automatic grouping procedure (EPICON).

[FIGURE 2]

[FIGURE 3]

[FIGURE 4]

[FIGURE 5]

[FIGURE 6]

[FIGURE 7]

[FIGURE 8]

[FIGURE 9]

[FIGURE 9]

[FIGURE 10]

[FIGURE 11]

[FIGURE 12]

[FIGURE 13]

THE EFFECT OF CASE-BASED REASONING

The comparison between morbidity rates based on EPI-0 (Figs. 6-9) and morbidity rates based on EPICON (Figs. 2-5) reveals the effect of case-based reasoning. Morbidity rates based on EPI-0 display a systematic bias, which manifests itself differently for prevalence and incidence rates.

Prevalence rates based on EPI-0 (Figs.6-7) show a systematic bias toward higher prevalence rates. EPI-0 produces obviously more episodes than the gold standard, which was to be expected since it processes all remaining diagnoses as separate episodes. For instance, EPI-0 always groups the remaining diagnosis 'other complaints of urine' (U07) as a separate episode, whereas medical coders, i.e., the gold standard, frequently grouped this diagnosis into an episode called 'cystitis/other urinary infection' (U71). The resulting prevalence rates for 'other complaints of urine' (U07) are 1.43 per 1000 patients based on EPI-0 (492 patients/ 343,853), and 1.02 per 1000 patients (350 patients/343,853) according to the gold standard, which is a difference of 40% (Fig. 7).

Incidence rates based on EPI-0 (Figs.8-9) show a systematic bias toward higher incidence rates for symptoms and complaints (codes 1-29), and lower incidence rates for diseases (codes 70-99). The total number of new episodes in EPI-0 is identical to the total number of new episodes in the gold standard; in both episode constructions, a new diagnosis is grouped as a separate episode. The episode name, however, can be different. For example, all new diagnoses 'other complaints of urine' (U07) were grouped as new episodes in EPI-0, whereas the gold standard grouped part of these diagnoses into new episodes of 'cystitis/ other urinary infection' (U71).

When case-based reasoning is added to the grouping system (EPICON), this systematic bias is removed and the precision of the estimates is improved.

Performance of EPICON in Split-sample Procedure

Figures 10-13 show the results of the splitsample method. Morbidity rates of the test set, that are based on the training set, show no systematic bias, but deviate more from the gold standard than the morbidity rates that are based on the full dataset. Figures of morbidity rates of the training set that are based on that same training set show a similar pattern (not shown). We therefore believe that a major part of the imprecision of our test set is due to inefficiency (for the training set is only half of the total dataset) and a minor part is due to overestimation.

DISCUSSION AND CONCLUSIONS

Results from this evaluation study indicate that EPICON is a useful tool to disclose data from EMRs in general practice for estimating morbidity rates. In general, morbidity rates based on EPICON deviate only slightly from the gold standard and show no systematic bias. EPICON performs less well, however, when it comes to diagnoses that can be linked to many other diagnoses, such as 'generalized/ unspecified pain' (A01).

This judgment is related to the purpose of EPICON, which is to estimate morbidity rates in general practice. These rates are used at the population level. For instance, to describe the extent of health problems, to forecast public health or to study health differences in time or between regions. Considering what morbidity rates are used for, some deviation from the gold standard, such as an increase from 13 to 14 per 1000 patients per year, is regarded as acceptable. Our judgment might be completely different with another purpose in mind. For example, we would not recommend using EPICON for a decision about the individual treatment of a patient.

This study revealed that the role of casebased reasoning in EPICON is very important for acquiring precise and unbiased morbidity rates. We started with a simple, not case-based method (EPI-0) to automatically group diagnoses into episodes. In the next step, we added

case-based reasoning, which led to the creation of EPICON. The evaluation showed that the performance of the grouping system improved considerably by adding case-based reasoning. Furthermore, the results indicate that the current set of cases is large enough for the purpose of estimating morbidity rates. Nevertheless, the performance of EPICON could be improved by increasing the number of cases, in particular by adding cases of patients with rare symptoms and diseases.

A limitation of our study is that both EPICON and the evaluation are based on the assumption that the standard is gold. This assumption seems plausible since data from the DNSGP-2 are widely used in the Netherlands, but there is no evidence to support this assumption. However, an evaluation of the gold standard was not a focus of our project. Therefore, this study cannot reveal possible flaws in the gold standard that were subsequently passed on to EPICON.

An evaluation of the gold standard would include a comparison between morbidity rates from the DNSGP-2 and rates from other registries in general practice.

Such a comparison is challenging, because the existing registries differ in factors as region of residence, patient population, classification system used to code diagnoses, measures taken to ensure the quality of the registration, practice software, and in the methods used to calculate morbidity rates [3]. More research is needed to quantify the impact of these factors on morbidity rates.

Findings from the split-sample procedure indicate that EPICON performs reasonably well on a simulated, independent dataset. Therefore, EPICON seems transportable to other similar datasets of diagnoses in general practice. The split-sample procedure is a tough test, because only half of the total dataset is used to construct the case bases. Less demanding tests [16] would probably show that overestimation is only a limited issue in this case.

We conclude that the internal validity of EPICON is satisfactory. EPICON seems useful for LINH and similar registries in general practice for the purpose of estimating morbidity rates. Further research should aim at the external validation of EPICON.

[FIGURE 14]

APPENDIX

Figure 14 shows EPICON in the form of a flowchart. EPICON consists of five steps (shown in between parentheses).

Step 1: The first consultation diagnosis of a patient in a one-year registration period is grouped as a separate episode (i.e. the diagnosis receives a separate episode number). Operationally, an episode is a row of diagnoses with the same episode number.

The first **diagnosis** of this row can be either new or ongoing, whereas subsequent diagnoses are always ongoing. The first diagnosis in the row determines whether the **episode** is new or ongoing, the last disease diagnosis in the row determines the name of the episode. If no disease diagnosis exists, the episode is named after the last symptom diagnosis in the row.

Step 2: The remaining diagnoses are grouped by ‘type of episode’ (new or ongoing).

A new diagnosis is always grouped as a separate episode.

Step 3: If possible, ongoing diagnoses are grouped by a decision table, which consists of a combination of an ongoing diagnosis and a previous diagnosis, and a decision whether or not they should be grouped together.

Step 4: The remaining diagnoses are grouped by case-based reasoning, which is a problem-solving approach in the field of artificial intelligence. A case-based reasoner solves a problem by remembering a previous similar situation and reuses information and knowledge from that situation. We used the manual grouping in the DNSGP-2 as a case library from which previous cases could be retrieved to solve the problem of grouping diagnoses.

EPICON uses three kinds of cases, one at the level of diagnoses (4a), one at the level of episodes (4b), and one at the level of chapters (4c).

In step 4a, an unsolved case is defined as: Should diagnosis X be linked to one of the other **diagnoses** of the same patient? A solved case is defined as a manually grouped diagnosis X that belongs to a patient who also has one of the other diagnoses.

EPICON reuses these solved cases to group diagnosis X.

In step 4b, EPICON uses the same procedure as in step 4a, but it uses a more coarse definition of cases in order to group any remaining diagnoses. Here, an unsolved case is defined as: Should diagnosis X be linked to one of the other **episodes** of the same patient? In step 4c, the grain size of cases is increased once more, to the level of chapters, to handle rare cases that cannot be grouped by any of the preceding steps. In this step, an unsolved case is described as whether diagnosis X should be linked to one of the **chapters** to which other diagnoses of the same patient belong.

Diagnoses that cannot be grouped by any of the aforementioned steps are set aside in step 4d. EPICON continuously tries to group these remaining diagnoses, because it might still be possible to group some of these diagnoses when more information from following diagnoses of the same patient becomes available. In the end, EPICON groups all remaining diagnoses as separate episodes.

Step 5: Diagnoses from prescriptions and referrals (i.e., the indications for the prescription or referral) do not pass through the foregoing steps, because they have no 'type of episode'. Diagnoses from prescriptions and referrals that are different from any of the consultation diagnoses of a patient are added to the database as ongoing episodes.

EPI-0 consists of the same steps, except for the case-based part (4a-4c), which does not exist in EPI-0.

REFERENCES

1. Boerma WG. Profiles of general practice in Europe (dissertation). Nivel; 2003.
2. Linden M, Gothe H, Ormel J. Pathways to care and psychological problems of general practice patients in a "gate keeper" and an "open access" health care system. A comparison of Germany and the Netherlands. Soc Psychiatry Psychiatr Epidemiol 2003; 38: 690-697.
3. Gijsen R, Poos MJJC. Using registries in general practice to estimate countrywide morbidity in the Netherlands. Public Health 2006; 120: 923-936.
4. WONCA Classification Committee. An international glossary for general/family practice. Fam Pract 1995; 12 (3): 341-369.
5. Biermans MCJ, De Bakker DH, Verheij RA, Gravestein JV, Van der Linden MW, De Vries Robbé PF. Development of a case-based system for grouping diagnoses in general practice. Int J Med Inform Epub 2007 Sep. 14.
6. Westert GP, Jabaaij L, Schellevis FG (editors).
Morbidity, performance and quality in primary care: Dutch general practice on stage. Oxon: Radcliffe; 2006.
7. van der Linden M, Westert GP, De Bakker DH, Schellevis FG. The second Dutch national survey of general practice. Health and problems in the Dutch population and in general practice (Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Klachten en aandoeningen in de bevolking en in de huisartspraktijk). Utrecht, Bilthoven: NIVEL, RIVM, 2004. Utrecht: Nivel; 2004.
8. Westert GP, Schellevis FG, de Bakker DH, Groenewegen PP, Bensing JM, van der Zee J.
Monitoring health inequalities through general practice: the Second Dutch National Survey of General Practice. Eur J Public Health 2005; 15 (1): 59-65.
9. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information.
Ann Intern Med 1999; 130: 515-524.
10. Tacken MAJB. Quality of preventive performance in general practice: the use of routinely collected data (dissertation). Radboud University Nijmegen; 2005.

11. Verheij RA, Jabaaij L, Abrahamse H, Van den Hoogen H, Braspenning J, Van Althuis T. Netherlands Information Network of General Practice.
Facts and figures on Dutch GP care (website on the internet, updated 2006, Aug 1; cited 2006, Nov 30). Available from: http://www.linh.nl 12. Lamberts H, Woods M, editors. International Classification of Primary Care (ICPC). Oxford: Oxford University Press; 1987.
13. Kenens R, Hingstman L. Figures from the registration of general practitioners. (Cijfers uit de registratie van huisartsen). Utrecht: Nivel; 2001.
14. Statistics Netherlands (database on the internet).
Voorburg/Heerlen: Statistics Netherlands. (Updated daily; cited 2005 Oct 14.) Available from: http://www.cbs.nl 15. Altman DG, Royston P. What do we mean by validating a prognostic model? Statist Med 2000; 19: 453-473.
16. Steyerberg EW, Harrell Jr FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. J Clin Epidemiol 2001; 54: 774-781.

FIGURES

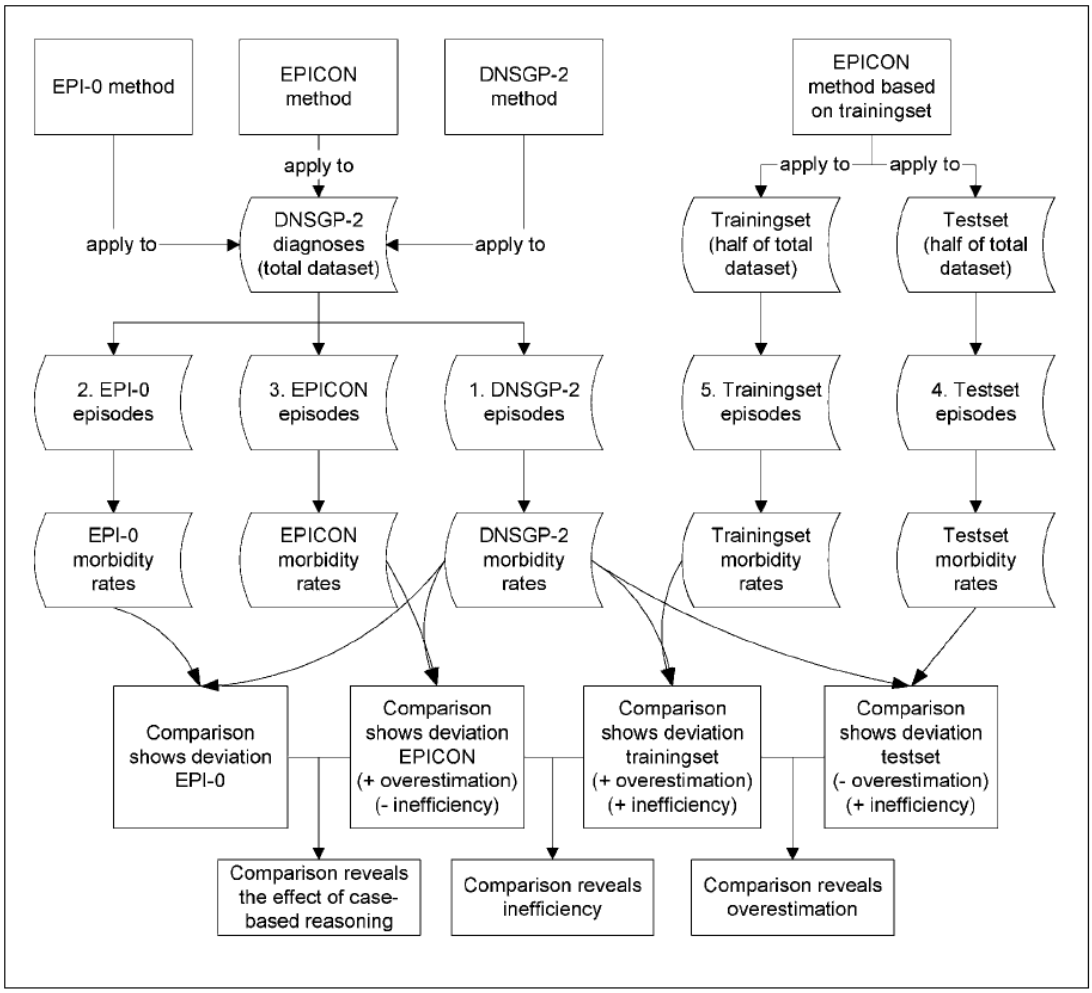

Fig. 1 Design of the study

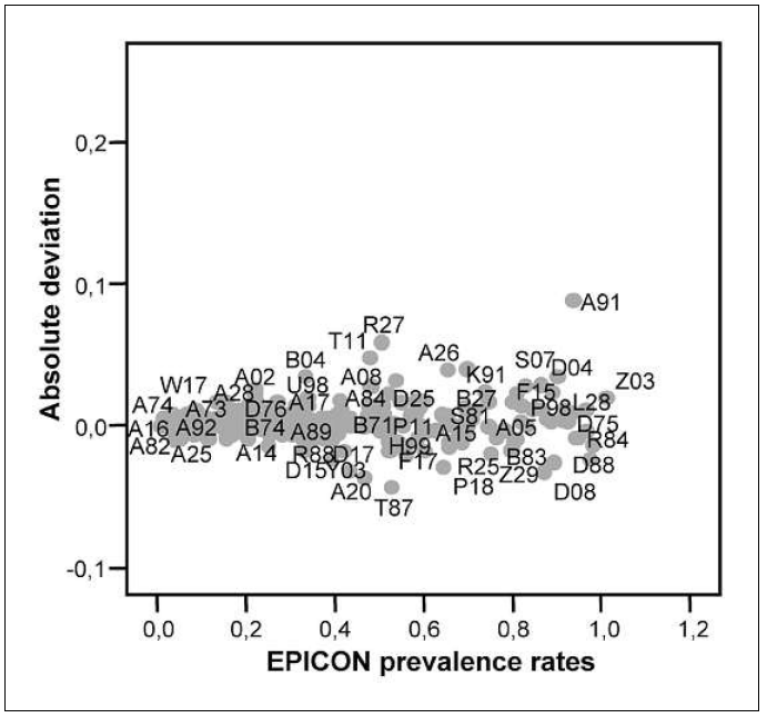

Fig. 2 Deviation from the gold standard for infrequent EPICON prevalence rates

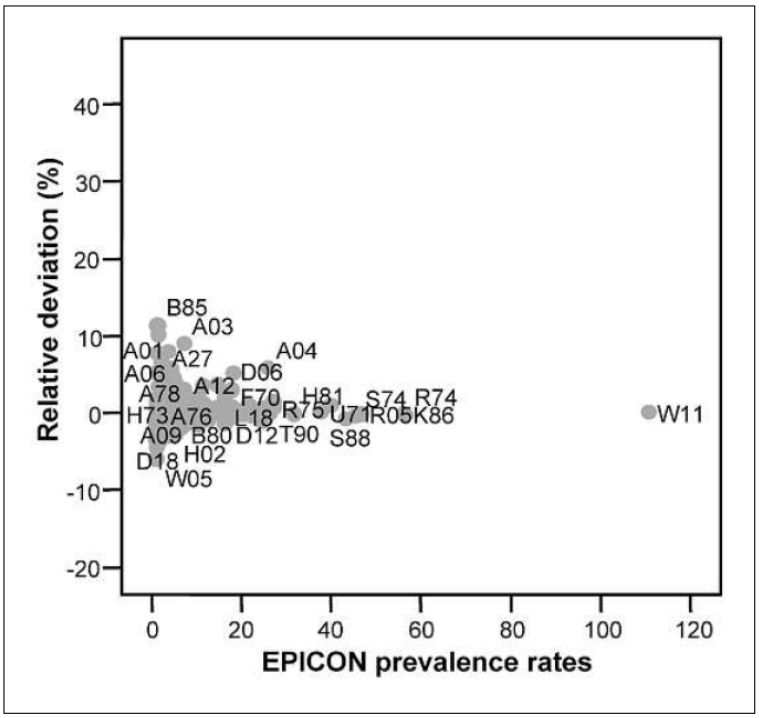

Fig. 3 Deviation from the gold standard for frequent EPICON prevalence rates

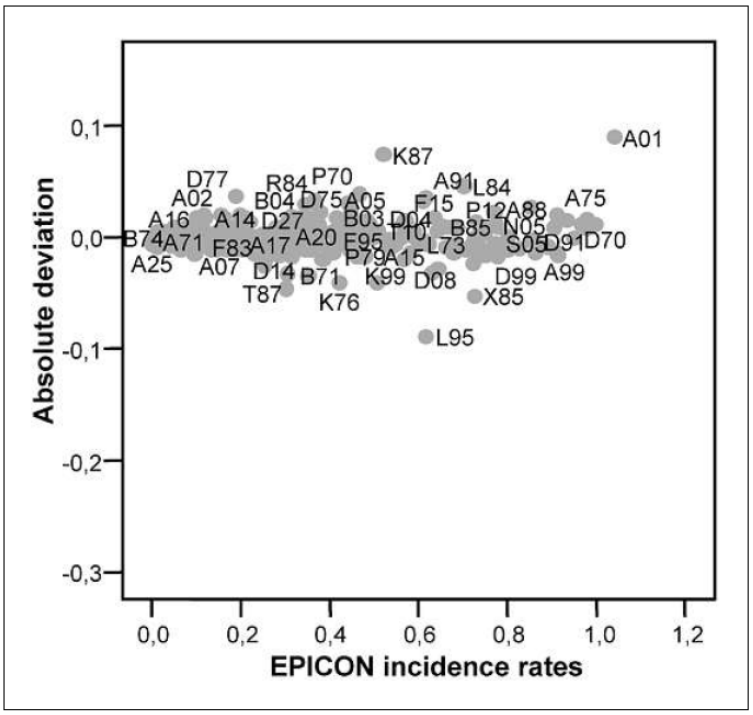

Fig. 4 Deviation from the gold standard for infrequent EPICON incidence rates

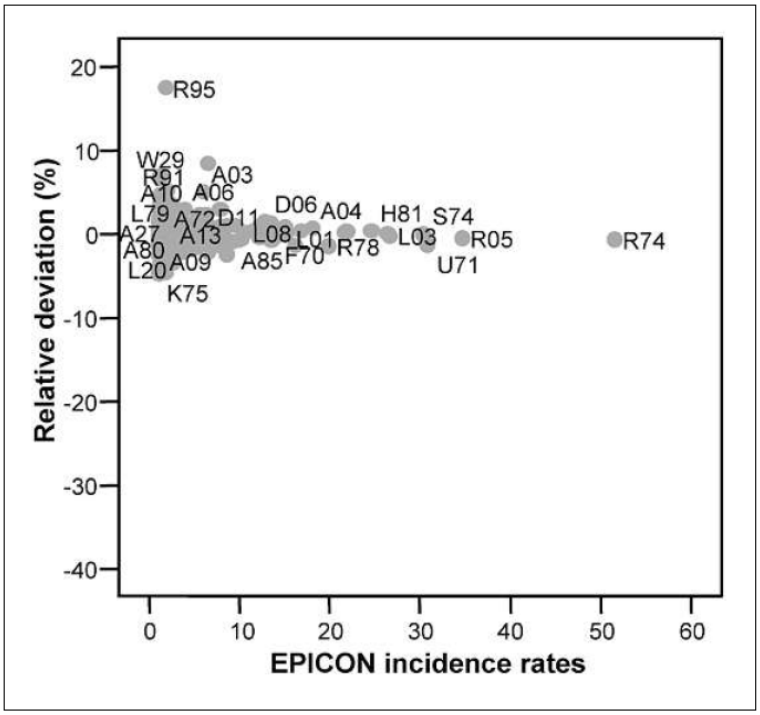

Fig. 5 Deviation from the gold standard for frequent EPICON incidence rates

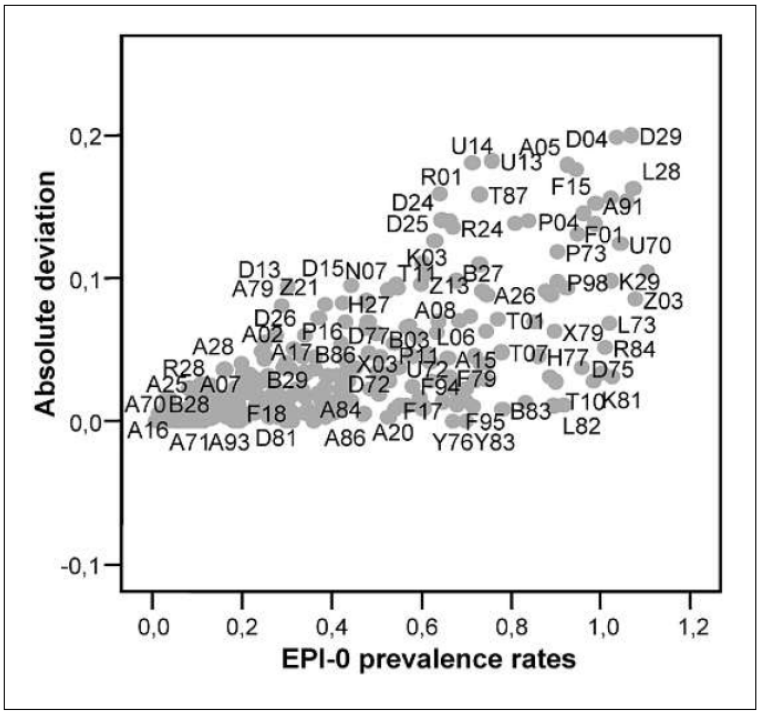

Fig. 6 Deviation from the gold standard for infrequent EPI-0 prevalence rates

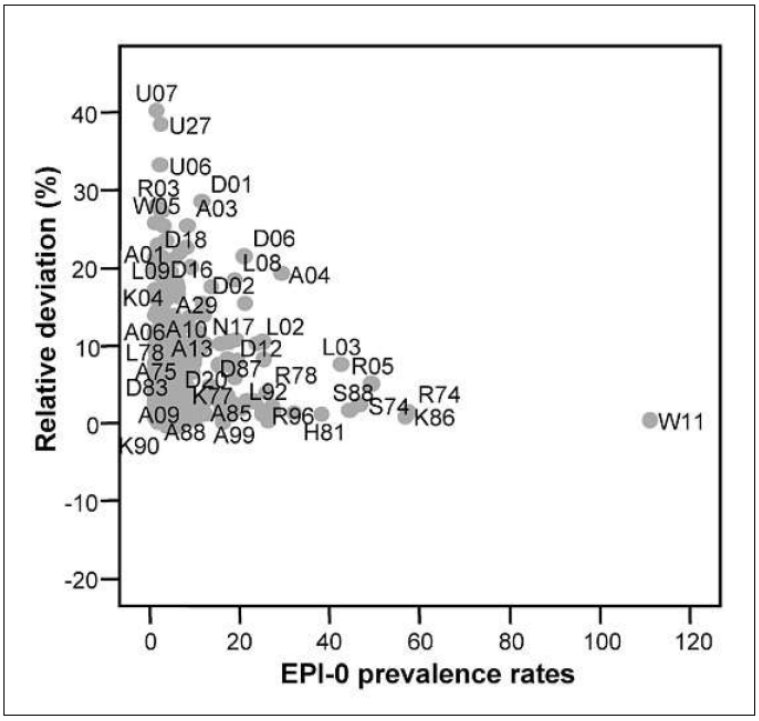

Fig. 7 Deviation from the gold standard for frequent EPI-0 prevalence rates

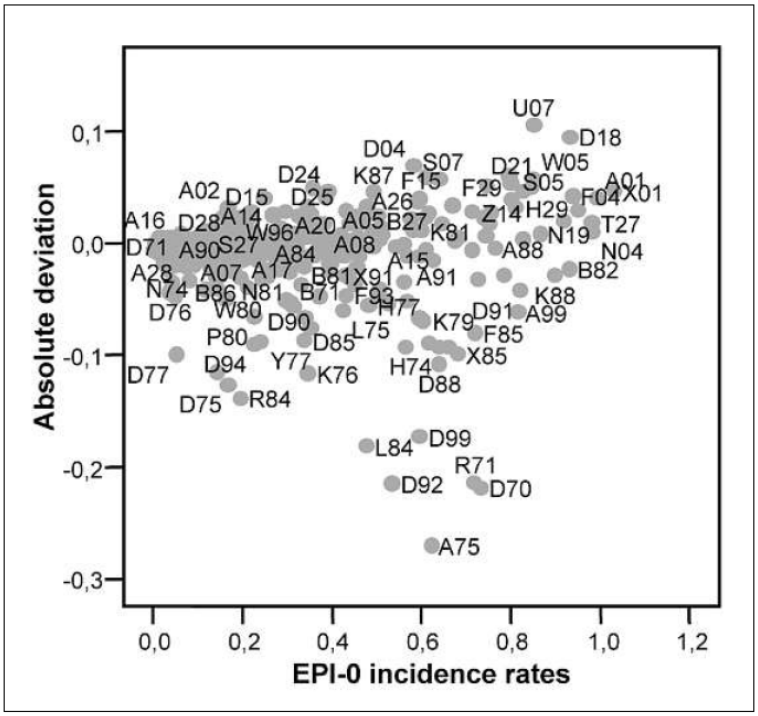

Fig. 8 Deviation from the gold standard for infrequent EPI-0 incidence rates

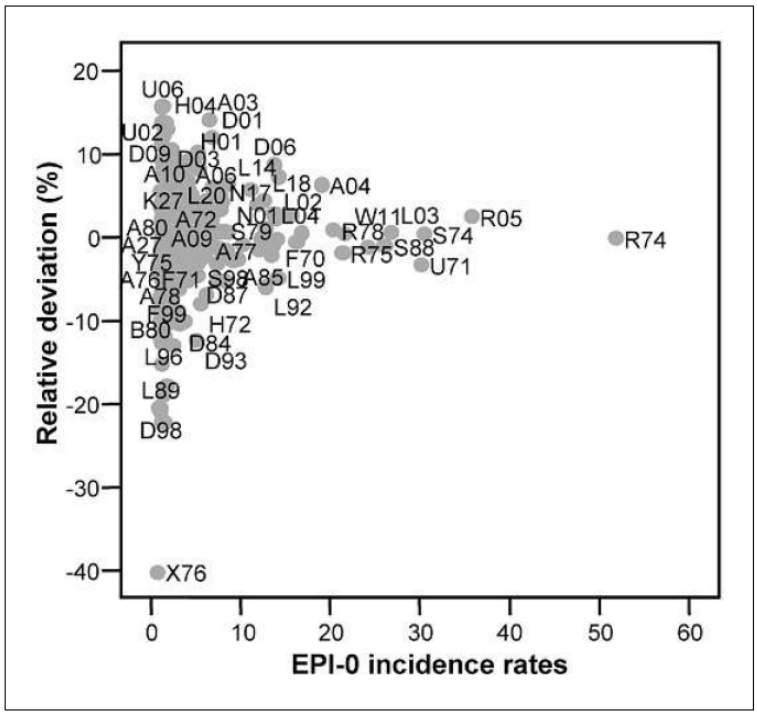

Fig. 9 Deviation from the gold standard for frequent EPI-0 incidence rates

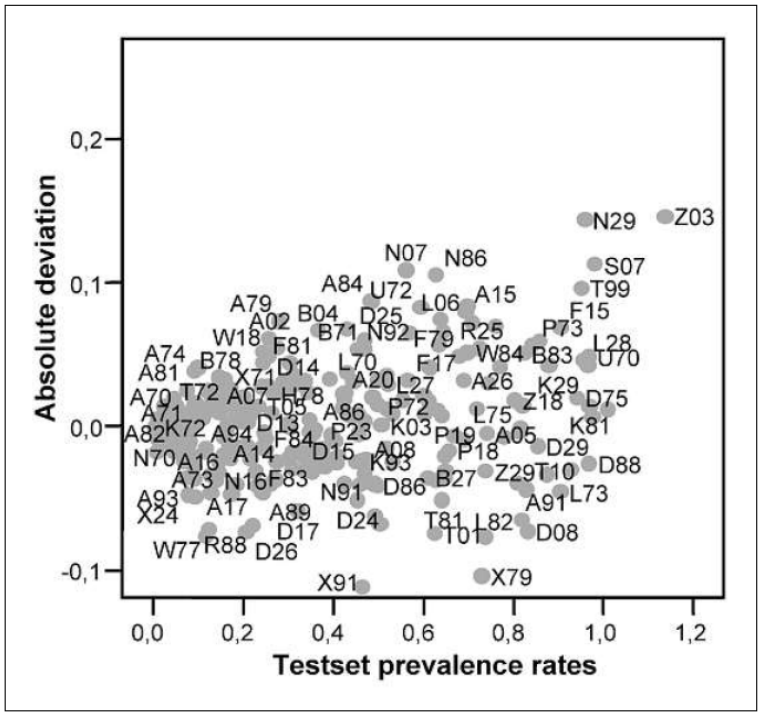

Fig. 10 Deviation from the gold standard for infrequent test set prevalence rates

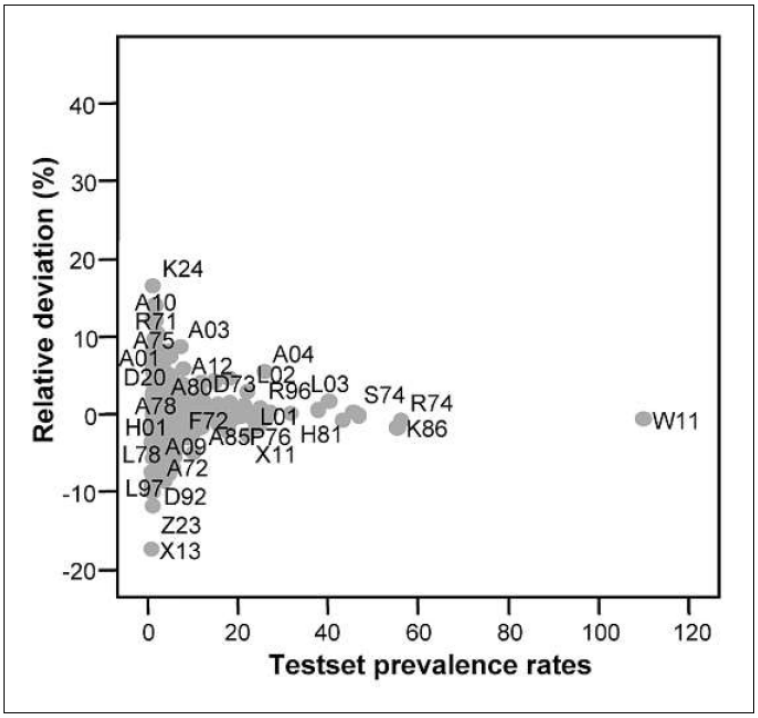

Fig. 11 Deviation from the gold standard for frequent test set prevalence rates

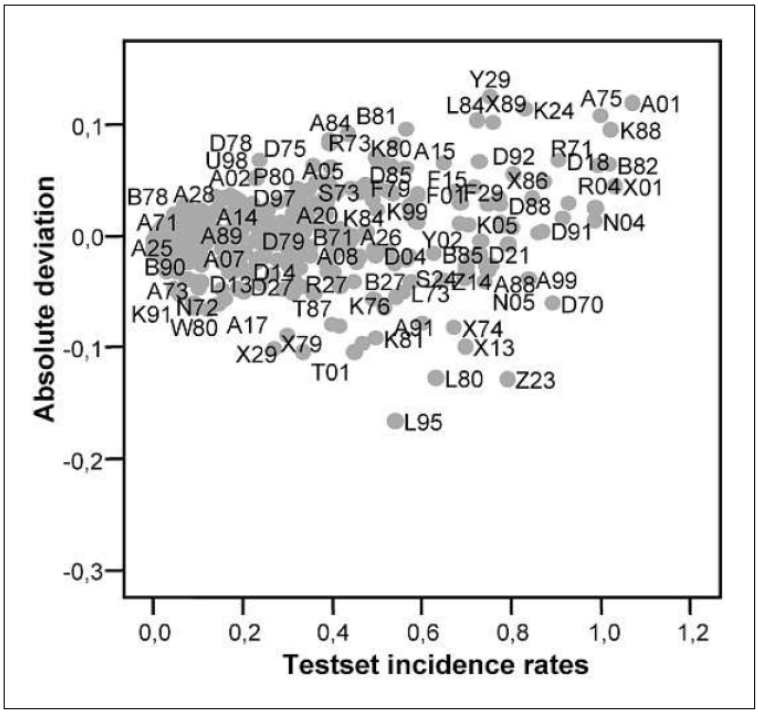

Fig. 12 Deviation from the gold standard for infrequent test set incidence rates

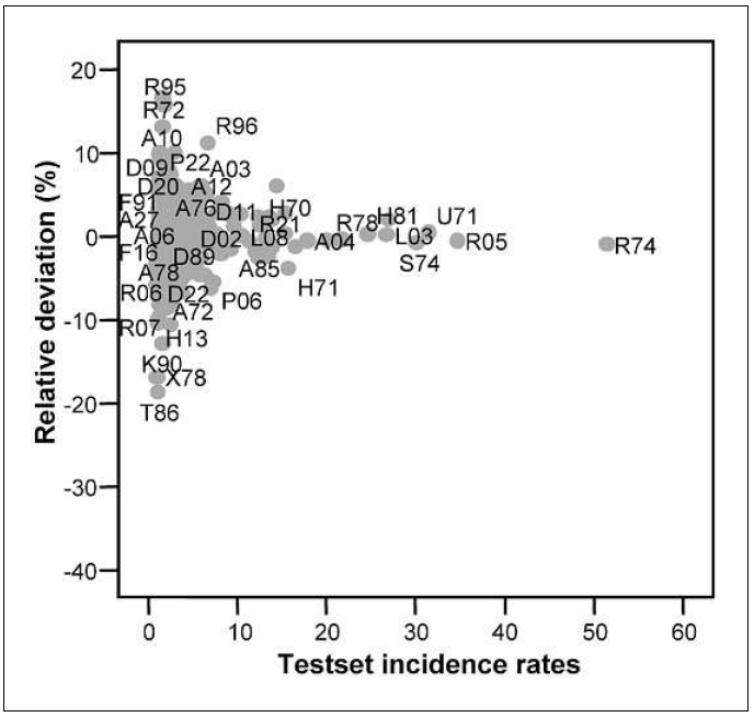

Fig. 13 Deviation from the gold standard for frequent test set incidence rates

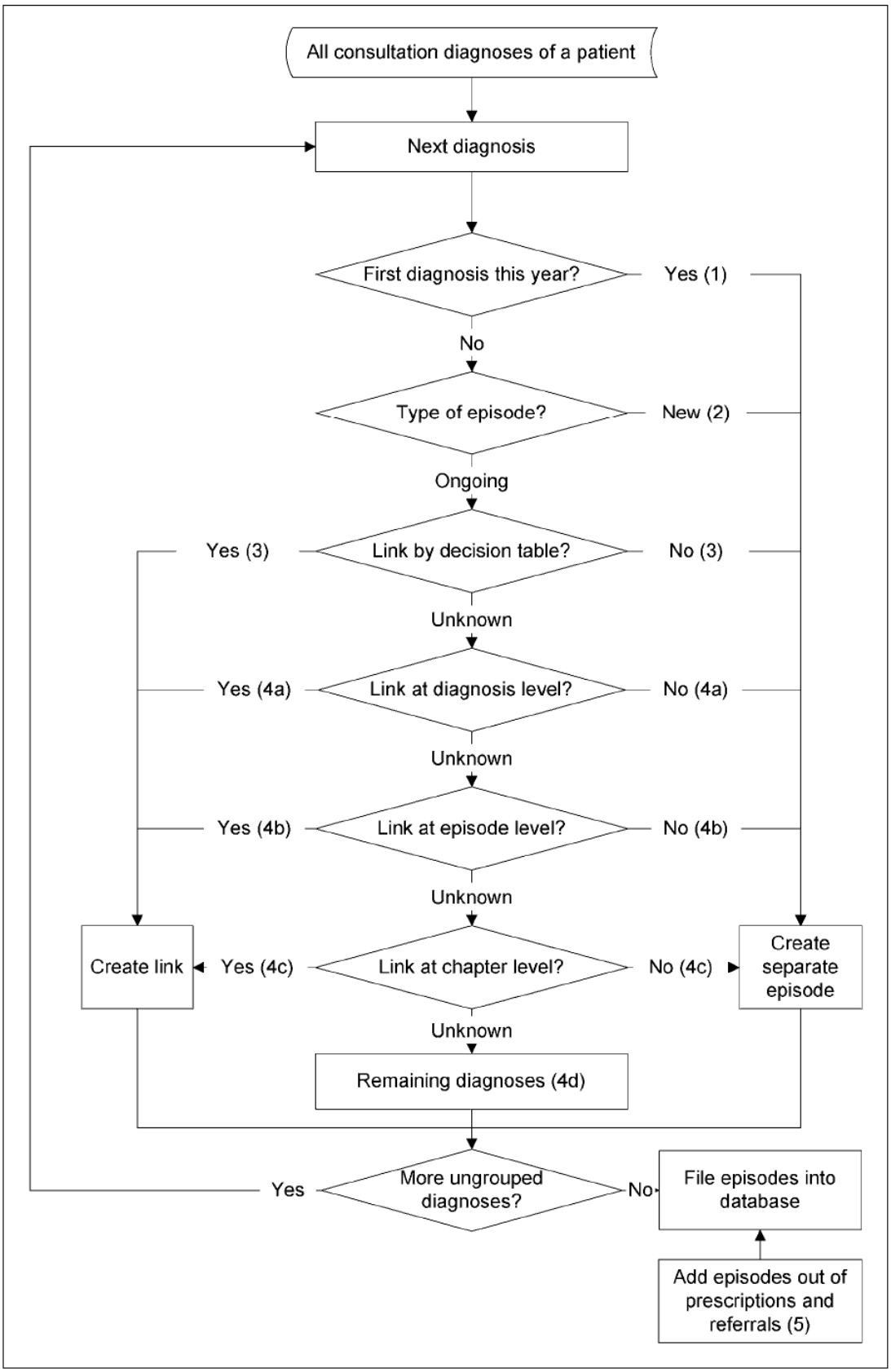

Fig. 14 EPICON