Validity of a procedure to identify patients with chronic idiopathic thrombocytopenic purpura in the Danish National Registry of Patients

Katrine Edith Klith Heden
Annette Østergaard Jensen
Dora Körmendiné Farkas
Mette Nørgaard

Department of Clinical Epidemiology,
Aarhus University Hospital, Denmark

Background: Administrative data may be useful for epidemiological studies of chronic idiopathic thrombocytopenic purpura (ITP). However, the quality of the recorded diagnoses needs evaluation.

Aim: We evaluated the validity in predicting chronic ITP of the International Classification of Diseases (ICD)-10 diagnoses of ITP in the Danish National Registry of Patients (NRP).

Methods: We used the NRP to identify patients with ITP, according to code D69.3, from January 1, 1996 to December 31, 2007. We defined chronic ITP as lasting longer than 6 months by including only patients with 2 or more hospital ITP diagnoses over longer than 6 months. We confirmed diagnoses by evaluating each candidate chronic ITP patient's medical chart and estimating the positive predictive value (PPV) and 95% confidence interval (CI) of the recorded NRP diagnostic code.

Results: We identified 513 patients with chronic ITP in the NRP. We were able to retrieve the charts of 439. After evaluation of the charts, 410 patients were deemed to have a valid diagnosis of chronic ITP, yielding a PPV of 0.93 (95% CI: 0.91–0.96).

Conclusion: The validity of this procedure to identify chronic ITP patients was high. The NRP is valid for epidemiological studies of patients with chronic ITP.

Keywords: validity, idiopathic thrombocytopenic purpura, National Registry of Patients

Introduction

Idiopathic thrombocytopenic purpura (ITP) is a disease characterized by low platelet counts and bleeding complications.[1] ITP patients develop autoantibodies, which cause premature platelet destruction and inhibit megakaryocyte production.[1,2] The shortened lifespan of platelets in the circulation, and the incomplete compensation from increased platelet production by bone marrow megakaryocytes, decreases the platelet count. Most frequently, ITP is followed by disease remission within 6 months. However, approximately 25% of ITP patients have persistent, chronic disease. Some patients are followed as long as 30 years without remissions.[3] Although chronic ITP is a common acquired bleeding disorder with potentially fatal outcomes, few data have been published on the epidemiology of this disease.

Medical registries, including all cases of a particular health condition or health-related exposure, are important data sources for epidemiological and clinical research.[4] The fact that the data already exist markedly reduces the cost and the time spent on such studies. Another advantage is that medical databases contain information on large study populations, offering the potential for precise estimates of effects and for studying rare exposures or outcomes. At the same time, the risk of bias, such as recall bias and non-response, is reduced. A shortcoming is that the researchers do not control the data selection and quality,[4,5] and that medical registries do not provide detailed clinical information.

Correspondence: Annette Østergaard Jensen
Department of Clinical Epidemiology, Aarhus University Hospital, Forskningens Hus, Sdr. Skovvej 15, DK-9000 Aalborg, Denmark
Tel +45 9932 6906
Fax +9932 6914
Email aoj@dce.au.dk

The Danish National Registry of Patients (NRP) is a medical registry. The validity of diagnoses in such medical registries may vary. Hardly any data exist on the predictive quality of two or more registry-recorded ITP diagnoses over a period of 6 months as a measure of chronic ITP. We therefore estimated the validity of recorded diagnoses of chronic ITP in the NRP from January 1, 1996 through 2007 by comparing the recorded diagnoses with data from chart reviews.

Methods

We conducted this population based cohort study in Denmark, which has a population of about 5.4 million inhabitants. The entire Danish population is provided tax-supported health care by the National Health Service, allowing free access to hospital care. Specialized medical centres within public hospitals, operating under the auspices of the Danish National Health Service, provide care of all ITP patients. We used the unique 10-digit civil registry numbers (CPR-numbers) assigned to each Danish resident[6] to link the national registry data of hospital diagnoses to the patients' medical charts.

Using the NRP in Denmark, we identified all adult (aged $\geq$ 18 years) patients with a hospital diagnosis of ITP from January 1, 1996 to December 31, 2007. The NRP, established in 1977, covers 99.4% of all patients admitted to general hospitals in Denmark.[7] Since 1995, all outpatient and emergency room visits were also included in the NRP. For each hospital visit, the NRP files include dates of admission and discharge, surgical procedure(s) performed, and up to 20 discharge diagnoses. Since 1994, diagnoses have been coded according to the 10th revision of the International Classification of Diseases (ICD-10).[7] We identified all patients registered with a hospital diagnosis of ITP (ICD-10 code D69.3). We further limited our study cohort to patients with chronic (lasting longer than 6 months) ITP by including only those with two or more hospital diagnoses of ITP over a period longer than 6 months.

We used the information from the chart reviews as a reference. In addition to requiring an ITP diagnosis recorded in the medical charts for a period of more than 6 months, we required a platelet count below $150 \times 10^9/L$ without an obvious cause to confirm the diagnosis. Patients who did not fulfil this criterion (eg, patients with secondary ITP, and patients who already were in remission for their ITP) were categorized as false positive in our study.

We estimated the positive predictive value (PPV) and associated 95% confidence intervals (CI) for a diagnosis of chronic ITP in the NRP compared with the results from the medical charts. The PPV was the proportion of patients registered with chronic ITP in the NRP, who also qualified as having chronic ITP according to our review of their charts. The numerator in this ratio was the number of patients registered with chronic ITP in both the NRP and the chart reviews, and the denominator was the number of patients with chronic ITP registered in the NRP. To quantify how potential misclassification among the patients for whom we were unable to retrieve the medical charts could affect the PPV, we repeated the analysis assuming a worst case scenario (ie, that all of the missing medical files did not fulfil the criteria for chronic ITP). We further estimated the impact of misclassification on relative risk estimates, assuming that the Danish NRP was used to identify chronic ITP patients in a cohort- and a case-control design. In these analyses, we assumed the misclassification in the chronic ITP diagnosis to be non-differential according to the exposure and outcome.

This study was approved by the Danish Data Protection Agency (Record no. 2007-41-1101). The statistical software package SAS, version 9.1 (SAS Institute Inc., Cary, NC), was used for all statistical analyses.

Results

Initially, we identified 2908 adult (age $\geq$ 18 years) patients registered in the NRP with a relevant ICD-10 code. Of these, 513 (17.6%) fulfilled our definition of chronic ITP (ie, an ITP diagnosis over the course of more than 6 months). The remaining 2395 patients (82%) were considered to have acute ITP. We were able to retrieve the medical charts on 439 (86%) of these patients. We carefully reviewed these medical charts, and found that 410 patients fulfilled the criterion of a chronic ITP diagnosis, yielding a PPV of 0.93 (95% CI: 0.91–0.96). The remaining 29 patients were categorized as false positive. If the 74 patients, for whom we were unable to retrieve the medical charts, did not fulfil the criterion of a chronic ITP diagnosis (worst case scenario), we would have obtained a PPV of 0.80 (95% CI: 0.76–0.83).

Among the 439 patients, the median age at the date of diagnosis was 56 years (range: 18–100 years). A total of 274 (62%) were women, and 165 (38%) were men. A total of 178 (41%) patients were diagnosed during the first 6 years of our study period (from 1996 through 2001), and 261 (59%) patients were diagnosed during the last 6 years (from 2002 through 2007). We found no difference in PPV according to gender or period of diagnosis (data not shown).

To estimate the impact of misclassification on relative risk estimates if the Danish NRP were used to identify chronic ITP patients in a cohort study, we simulated a study cohort

of 400 chronic ITP patients and 400 persons without chronic ITP. The chronic ITP patients and persons without chronic ITP were followed until death in a period of five years. If the PPV of the chronic ITP diagnosis were 0.93, a correct relative risk (RR) for death (eg, mortality) among chronic ITP patients compared with persons without chronic ITP of 2.00 would decline to 1.94. In a worst-case scenario, a correct RR of 2.00 would decline to 1.78, if the misclassification of the chronic ITP diagnosis were 0.80. We likewise simulated a case-control study of 400 cases with chronic ITP and 400 controls without chronic ITP, and estimated the odds ratio for chronic ITP if cases and controls had a history of a given exposure. A correct odds ratio (OR) of 2.11 would decline to 2.04, if the misclassification of the chronic ITP diagnosis were 0.93. In a worst-case scenario, a correct OR of 2.00 would decline to 1.88. For further details see Table 1.

Discussion

Using our procedure, comparing the data quality in the NRP with information obtained from chart reviews, we found that the predictive value of NRP diagnoses for identifying chronic ITP patients was 93%. This is above the accuracy of most validated diagnoses in the NRP.[8,9]

Our findings suggest that the NRP can be a valuable tool in register-based studies, bearing in mind that the necessary level of data quality depends on the research question and design. The positive predictive value should be stable over time, if NRP data are used to compare the incidence of chronic ITP over time. We found this to be the case within the 12-year time-period of our study. However, we do not know whether there were any time trends in the data quality before 1996. It is important that the small misclassification we found is unrelated to information about earlier exposures or future outcomes (ie, non-differential misclassification), if NRP data are used in case-control or cohort studies. We expect this to be true for most exposures and outcomes. Our analysis, illustrating the impact of misclassification on risk estimates if the NRP were to be used to identify chronic ITP patients for a cohort design or a case control design, shows that (under the assumption that the misclassification is non-differential according to exposure and outcome) the results would be biased toward conservative risk estimates. If the exposure had been misclassified differentially according to chronic ITP status, or if chronic ITP patients had been misclassified differentially according to their prognosis, the risk estimates would be exaggerated or underestimated.[10]

Table 1 The impact of misclassification on various risk estimates if the Danish National Patient Registry were used to identify chronic ITP (cITP) patients in a cohort study and a case-control study

	Correct classification		Positive predictive value 0.94		Positive predictive value 0.80 (worst case scenario)	
	cITP yes	cITP no[1]	cITP yes	cITP no	cITP yes	cITP no
Cohort study						
Deaths	10	5	9.3	5	8	5
Alive	390	395	362.7	395	320	395
False positive[2]	0	0	28	0	72	0
Mortality rate (%) after correction for misclassification	2.5	1.25	2.42	1.25	2.23	1.25
RR[3]	2.00		1.94		1.78	
Case-control study						
Exposed	40	20	37.6	20	32	20
Controls	360	380	338.4	380	288	380
False positive[4]	0	0	24	0	80	0
Number of exposed after correction for misclassification	40	20	38.8	20	36	20
Number of controls after correction for misclassification	360	380	361.2	380	364	380
OR[5]	2.11		2.04		1.88	

[1]Misclassification of the chronic ITP diagnosis will not affect the background population's risk of chronic ITP because the disease is rare.
[2]The misclassified patients will have the same mortality rate as the background population.
[3]The relative risk (RR) is independent of the study population's size.
[4]The misclassified patients will have the same exposure prevalence as the background population.
[5]The odds ratio (OR) will change towards the estimate of the RR, as the size of the study population increases.

Moreover, we used chart reviews for the validation of our procedure to identify patients with chronic ITP in the NRP, and we were able to estimate the PPV. Another dimension of validity is sensitivity and specificity. The only way we could identify chronic ITP patients was through our procedure, and we only used chart reviews for the validation of diagnoses identified through this procedure. Thus, we were unable to estimate sensitivity because our design did not allow us to identify patients who had chronic ITP recorded in their medical charts but did not have chronic ITP registered in their coding. Because we did not know the true incidence of chronic ITP in the background population, we were unable to estimate specificity.[4]

The diagnostic process of ITP is complex because it is, in large part, one of exclusion. Patients have thrombocytopenia, defined as a platelet count below 150×10^9/L without any obvious cause. Other causes of thrombocytopenia, such as disseminated intravascular coagulation, vitamin deficiency, infection, and primary bone marrow disease, need to be considered and ruled out before the diagnosis of ITP can be made. Furthermore, ITP is often associated with other autoimmune diseases, such as systemic erythematic lupus (SLE), diabetes mellitus, and morbus Basedow. Therefore, these also need to be ruled out. It is likely that the high validity of our procedure for identifying chronic ITP patients in the NRP is due to the complex diagnostic process. Moreover, we only included patients with more than one admission or outpatient visit registered with an ITP diagnosis over a period longer than 6 months. Such a diagnostic procedure lowers the probability of a false positive ITP diagnosis. Instead, we may have missed some patients with chronic ITP, for instance, those patients who were only admitted to hospital with their chronic ITP once, and then referred to their general practitioners for future control. However, we would expect this to happen only for patients with a relatively high baseline platelet count.

It is a major advantage to be able to use existing data sources with large amounts of information, such as the NRP, for research purposes. We found a high predictive value with our procedure for identifying chronic ITP patients in the NRP, suggesting that the NRP can be a valuable tool in clinical epidemiological studies of chronic ITP patients.

Acknowledgments

Disclosures

The authors have no conflicts of interest to disclose in relation to the present study.

References

1. Kuwana M, Kurata Y, Fujimura K, et al. Preliminary laboratory based diagnostic criteria for immune thrombocytopenic purpura: evaluation by multi-center prospective study. *J Thromb Haemost*. 2006;4(9):1936–1943.
2. McMillan R, Wang L, Tomer A, Nichol J, Pistillo J. Suppression of in vitro megakaryocyte production by antiplatelet autoantibodies from adult patients with chronic ITP. *Blood*. 2004;103(4):1364–1369.
3. Hirsch EO, Dameshek W. Idiopathic thrombocytopenia; review of eighty-nine cases with particular reference to the differentiation and treatment of acute (self-limited and chronic types). *AMA Arch Intern Med*. 1951;88(6):701–728.
4. Sorensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25(2):435–442.
5. Norgaard M, Skriver MV, Gregersen H, Pedersen G, Schonheyder HC, Sorensen HT. The data quality of haematological malignancy ICD-10 diagnoses in a population-based hospital discharge registry. *Eur J Cancer Prev*. 2005;14(3):201–206.
6. Frank L. Epidemiology. When an entire country is a cohort. *Science*. 2000;287(5462):2398–2399.
7. Andersen TF, Madsen M, Jorgensen J, Mellemkjoer L, Olsen JH. The Danish National Hospital Register. A valuable source of data for modern health sciences. *Dan Med Bull*. 1999;46(3):263–268.
8. Mosbech J, Jorgensen J, Madsen M, Rostgaard K, Thornberg K, Poulsen TD. [The national patient registry. Evaluation of data quality]. *Ugeskr Laeger*. 1995;157(26):3741–3745.
9. Nickelsen TN. [Data validity and coverage in the Danish National Health Registry. A literature review]. *Ugeskr Laeger*. 2001;164(1):33–37.
10. Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology*. 2007;18(1):67–72.