

Validity of asthma diagnoses in the Danish National Registry of Patients, including an assessment of impact of misclassification on risk estimates in an actual dataset

Annette Østergaard Jensen¹
Gunnar Lauge Nielsen²
Vera Ehrenstein¹

¹Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; ²Department of Medicine, Himmerland Hospital, Farsø, Denmark

Objective: Asthma diagnoses recorded in the Danish National Registry of Patients (DNRP) are a misclassified measure of the actual asthma status. We quantified this misclassification and examined its impact on the results of an epidemiologic study on asthma.

Study design and setting: We validated the DNRP asthma diagnoses against records of asthma diagnosed at medical examinations conducted during mandatory conscription evaluation. We had data on 22,177 male conscripts who were born from January 1st, 1977 to December 31st, 1983, in a conscription district in northern Denmark. We obtained asthma diagnoses recorded among the conscripts in the DNRP from January 1st, 1977 through December 31st, 2003. We estimated sensitivity, specificity, and positive predictive value (PPV) of the DNRP asthma diagnoses. We then conducted sensitivity analysis to quantify the impact of nondifferential misclassification on the rate ratios measuring the association between asthma and risks of different skin cancers.

Results: The sensitivity of the DNRP for detecting an asthma diagnosis was 0.44 (95% confidence interval [CI]: 0.42–0.47), the specificity was 0.98 (95% CI: 0.98–0.99) and the PPV was 0.65 (95% CI: 0.62–0.68). Both direct and inverse associations between asthma and the different types of skin cancers became more pronounced after correcting for the misclassification.

Conclusion: The DNRP registered asthma diagnosis may be used to measure asthma status in epidemiologic studies seeking to estimate relative effects of asthma. Even at low values of DNRP sensitivity of asthma diagnoses were not sufficient to nullify observed relative associations in an actual dataset. The specificity of DNRP asthma diagnosis is high.

Keywords: asthma, validity, registry data, epidemiology

Introduction

Asthma is a chronic disease that can present at any age, but the peak age of onset is around 6 years.¹ Asthma, with a prevalence of 10%, is subject of many epidemiologic studies, and ascertainment of asthma commonly relies on hospital records. In order to draw valid inferences from such studies, the validity of the asthma diagnosis in hospital records needs to be quantified.²

Medical registries and databases are increasingly used as data sources in epidemiologic research, including that of asthma. The Danish National Registry of Patients (DNRP) is a population electronic medical registry³ that records diagnoses made during hospital contacts. One study that used medical records as the gold standard to validate inpatient discharge diagnoses of asthma in the DNRP among children aged

Correspondence: Annette Ø Jensen
Department of Clinical Epidemiology,
Aarhus University Hospital, Olof Palmes
Allé 43–45, DK-8200
Aarhus N, Denmark
Tel +45 8942 4813
Fax +45 8942 4801
Email aoj@dce.au.dk

6–14 years, reported a sensitivity of 90%, a specificity of 99%, and a positive predictive value (PPV) of 85%.⁴ The validity of asthma diagnoses in the DNRP among adults has not been assessed, and neither has the validity of outpatient clinic contacts and emergency department records of asthma.

We estimated sensitivity, specificity, and PPV of a recorded diagnosis of asthma in the DNRP against asthma diagnoses recorded among men presenting for their mandatory medical evaluation at the Danish military draft board. We then used the estimated sensitivity and specificity to examine the impact of asthma status misclassification on estimates from a study of the relation between asthma and risk of skin cancers.

Materials and methods

Study base

We conducted this study in the fifth military conscription district of Denmark,⁵ with geographic jurisdiction primarily over former North Jutland and Viborg counties (population of approximately 700,000 inhabitants, or 10% of the total Danish population). We had data on 22,177 men who were born from January 1st, 1977 to December 31st, 1983 in the conscription district and presented before the draft board authorities from January 1st, 1995 to December 31st, 2003. This dataset, known as the North Jutland conscription dataset, has been maintained at the Department of Clinical Epidemiology, and used for various research projects.³

All men are draft liable in Denmark and are required by law to appear before the draft board at age 18 years for a mandatory medical evaluation to determine suitability to serve. Prior to actual appearance at the board, men receive a postal health questionnaire, in which they can report diseases potentially precluding army service. The draft board authorities verify such reports either with each conscript's healthcare provider or in an actual physical examination. All diagnoses established at conscription are recorded in the conscripts' files regardless of whether they lead to army rejection. The draft board used the *International Classification of Diseases*, 10th revision (ICD-10) to record the diagnoses during the study period. Because diagnoses at conscription are recorded after a formal verification or face to face contact with a physician, these asthma diagnoses at conscription were assumed to reflect the actual asthma status better than the DNRP and were therefore used as the gold standard for this study. A conscript was considered to have asthma if his file contained an ICD-10 diagnostic code J45.xx.

The Danish National Registry of Patients

Since 1977, Danish counties have maintained administrative information systems to monitor hospital admissions, collecting information on dates of admission and discharge and up to 20 discharge diagnoses. Data from these information systems are transferred to the DNRP, which contains data on 99.4% of all discharge records from Danish hospitals.⁶ From 1977 to 1993, discharge diagnoses have been coded according to the ICD-8, and thereafter according to the ICD-10.⁶ Since 1995, outpatient contacts and emergency room visits also have been recorded in addition to inpatient hospital stays. We obtained all asthma diagnoses among the conscripts recorded in the DNRP from January 1st, 1977 to December 31st, 2003.

Linkage of data sources

All public databases in Denmark uniquely identify persons by civil registration (CPR) number. This 10-digit number, encoding sex and date of birth, has been assigned at birth or emigration by the Danish Civil Registration System since 1968.⁷ We used the CPR number to link DNRP records with the conscription records.

The cohort study of asthma and risk of skin cancers

To estimate the impact of misclassification of asthma coded in the DNRP on relative estimates in a cohort study, we used results of our unpublished nationwide cohort study, assessing the risk of skin cancers among asthmatic patients. We examined risks of malignant melanoma (MM), basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), among 155,364 patients with a first time DNRP diagnosis of asthma recorded from 1977 through 2003. The asthma patients were followed for 1,168,944 person-years, while the skin cancers were ascertained by linkage to the Danish Cancer Registry. We calculated standardized incidence rate ratios (SIRR) for asthma, by dividing the case count among the asthmatics (observed) by the number of cases in the same amount of general population person time (expected). The observed SIRR were 0.87 (95% confidence intervals [CI]: 0.74–1.03) for MM (144 observed vs. 165 expected cases), 1.13 (95% CI: 1.05–1.20) for BCC (969 vs. 862), and 1.46 (95% CI: 1.23–1.67) for SCC (176 vs. 123).

Statistical analyses

Sensitivity of the DNRP diagnosis of asthma was calculated as the proportion of men with a record of asthma at conscription who also had an DNRP record of asthma (ICD-8 code

Table 1 General format for contingency tables

	Misclassified values			Corrected values		
	Asthma			Asthma		
	Yes	No	Total	Yes	No	Total
Skin cancer cases	A1*	A0*	A	A1	A0	A
Person-years (PY)	PY1*	PY0*	PY	PY1	PY0	PY
	$IRR^* = \frac{A1^*/PY1^*}{A0^*/PY0^*}$			$IRR = \frac{A1/PY1}{A0/PY0}$		

493.xx or ICD-10 code J45.xx). Specificity of the DNRP recorded asthma was calculated as the proportion of all men registered without record of asthma at conscription who also had no DNRP record of asthma. The PPV of a DNRP diagnosis of asthma was calculated as the proportion of all men registered with asthma in the DNRP, who also received asthma diagnosis at conscription. The PPV was calculated overall and by hospital contact type (inpatient stay, outpatient clinic contact, emergency department visits). The main purpose with the PPV was to see any differences between inpatient stay coding, outpatient clinic contact coding, and emergency department visit coding.

We then used these estimates of sensitivity and specificity to correct for misclassification estimates of relative associations in the study of asthma and skin cancers. By design, the person time of asthma exposed and asthma unexposed was artificially set to be equal in order to calculate SIRR. This does not reflect the actual distribution of person time contributed by asthma patients to the general population, which is needed for a sensitivity analysis of misclassification impact. Therefore, we constructed hypothetical contingency tables from which regular IRR could be calculated. Following Greenland,⁸ we used the following general format to construct contingency tables (Table 1).

To these misclassified IRR estimates, we applied the point estimates of sensitivity and specificity to calculate corrected cell entries and IRR for each of the skin cancer types.⁹ We assumed that the rate of misclassification of asthma status was the same for patients with and without skin cancer (ie, non-differential misclassification). We, then, estimated apparent IRR that would be observed for each of the outcomes under the different assumptions about sensitivity and specificity of the hospitalization records.¹⁰ In particular, we were interested in how severe a misclassification would produce an apparent absence of effect (ie, IRR = 1).

This study was approved by the Danish Data Protection Agency. The SAS statistical software package (version 9.2; SAS Institute Inc., Cary, NC) was used for all statistical analyses.

Results

Validity and predictive value of DNRP asthma diagnosis

Of the 1358 men with a diagnosis of asthma recorded at conscription, 604 men also had an asthma diagnosis recorded in the DNRP, resulting in a sensitivity of 0.44 (95% CI: 0.42–0.47). Of the 20,819 men who did not have a diagnosis of asthma recorded at conscription 20,498 also did not have a diagnosis recorded in the DNRP. This resulted in a specificity of 0.98 (95% CI: 0.98–0.99) (Table 2).

A total of 925 of these men had a diagnosis of asthma coded in the DNRP, and of these, 604 men were also recorded in the conscription dataset with a diagnosis of asthma, resulting in a PPV of 0.65 (95% CI: 0.62–0.68). A total of 621 of the 925 asthma diagnosis codes were recorded as inpatient stays, 289 were recorded as outpatient visits, and 15 were emergency department visits. The PPV for an inpatient stay with asthma diagnosis was 0.65 (95% CI: 0.62–0.68), for an outpatient clinic contact 0.65 (95% CI: 0.62–0.69), and for an emergency department visit for asthma 0.73 (95% CI: 0.51–0.96). For further details, see Table 3.

Consequences of application of observed validity on the cohort study of asthma and risk of skin cancers

Assuming the point estimates of the sensitivity and specificity of an asthma diagnosis in the DNRP (44% and 98%,

Table 2 Validity of the diagnosis codes of asthma in the DNPR, January 1st, 1977 to December 31st, 2003

Danish National Registry of Patients	Military Conscription Registry	
	With asthma diagnosis	Without asthma diagnosis
With asthma diagnosis	604	321
Without asthma diagnosis	754	20,498
Total	1,358	20,819
Sensitivity (95% CI):	604/1,358 = 0.44 (0.42–0.47)	
Specificity (95% CI):	20,498/20,819 = 0.98 (0.98–0.99)	

Abbreviations: CI, confidence intervals; DNRP, Danish National Registry of Patients.

Table 3 Positive predictive value of the diagnosis codes of asthma in the DNRP, January 1st, 1977 to December 31st, 2003 by hospital contact type

	Danish National Registry of Patients		
	Military conscription registry	With asthma diagnosis	Without asthma diagnosis
Asthma total	With asthma diagnosis	604	754
	Without asthma diagnosis	321	20,498
	Total	925	21,252
	PPV (95% CI):	604/925 = 0.65 (0.62–0.68)	
Asthma hospitalization (ie, inpatient stay)	With asthma diagnosis	406	754
	Without asthma diagnosis	215	20,498
	Total	621	21,252
	PPV (95% CI)	406/621 = 0.65 (0.62–0.69)	
Asthma outpatient clinic contact	With asthma diagnosis	187	754
	Without asthma diagnosis	102	20,498
	Total	289	21,252
	PPV (95% CI)	187/289 = 0.65 (0.59–0.70)	
Asthma emergency department visit	With asthma diagnosis	11	754
	Without asthma diagnosis	4	20,498
	Total	15	21,252
	PPV (95% CI)	11/15 = 0.73 (0.51–0.96)	

Abbreviations: CI, confidence interval; DNRP, Danish National Registry of Patients; PPV, positive predictive value.

respectively), a misclassified IRR of 0.87 for MM among asthma patients would decline to 0.79, if there was no asthma misclassification. Accordingly, a misclassified IRR of 1.13 for BCC among asthma patients would increase to 1.21, if there was no asthma misclassification. Finally, a misclassified IRR of 1.46 for SCC among asthma patients would increase

to 1.77, if there was no asthma misclassification. For further details, see Table 4A–C.

Figure 1A–C demonstrates the impact of misclassification of asthma diagnosis by the DNRP on estimates of IRR, using ranges of values of the sensitivity and specificity. In order to bring the apparent IRR to the value of 1 (ie, to produce

Table 4 A–C The impact of misclassification on various risk estimates if the DNPR were used to identify asthma patients' risk of skin cancers, misclassified and corrected for misclassification

	Misclassified values			Corrected values		
	Asthma			Asthma		
	Yes	No	Total	Yes	No	Total
A. Malignant melanoma						
Cases with asthma	6.00	159.00	165	7.85	157.15	165
Person-years at risk	48,744.96	1,120,199.04	1,168,944	69,847.59	1,099,096.41	1,168,944
	IRR* = 0.8679			IRR = 0.7862		
B. Basal cell carcinoma						
Cases with asthma	40.41	821.59	862	61.74	800.26	862
Person-years at risk	48,744.96	1,120,199.04	1,168,944	69,847.59	1,099,096.41	1,168,944
	IRR* = 1.1302			IRR = 1.2140		
C. Squamous cell carcinoma						
Cases with asthma	7.34	115.66	123	12.42		123
Person-years at risk	48,744.96	1,120,199.04	1,168,944	69,847.59	1,099,096.41	1,168,944
	IRR* = 1.4582			IRR = 1.7669		

Abbreviation: IRR, incidence rate ratio.

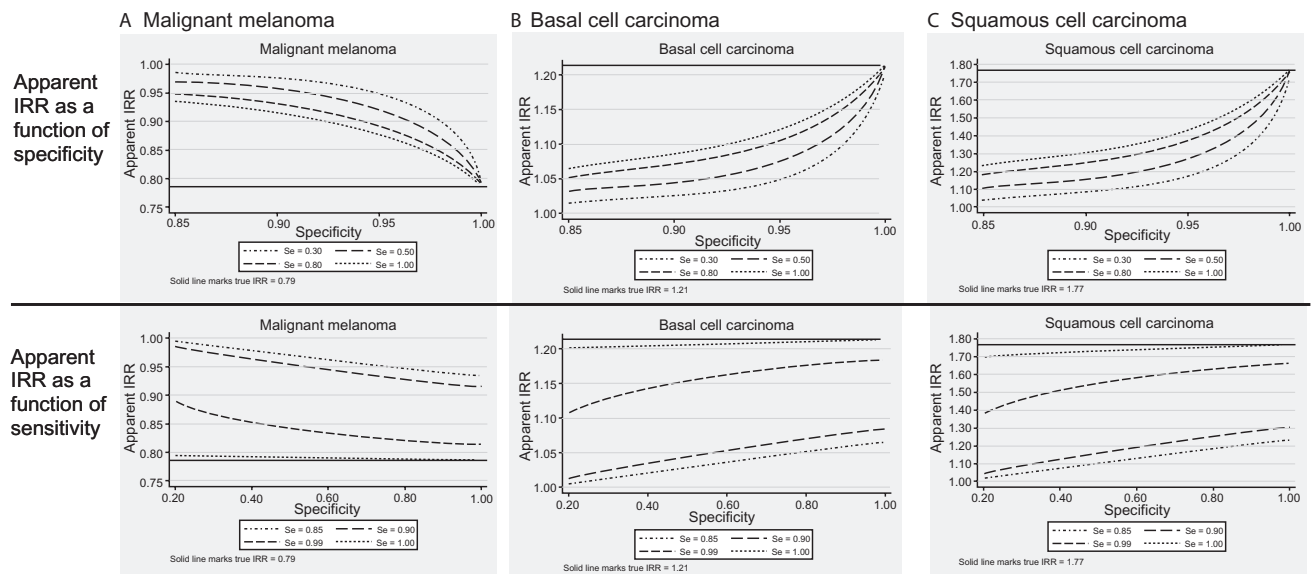


Figure 1A–C The impact of misclassification on various risk estimates if the Danish National Patient Registry were used to identify asthma patients' risk of skin cancers for varying values of sensitivity and specificity.

Abbreviation: IRR, incidence rate ratios.

an apparent absence of effect), sensitivity of the DNRP in detecting asthma would need to be as low as 20%, combined with a specificity of 85%–90%. These estimates are much lower than those estimated in our analyses.

Discussion

The sensitivity of a recorded diagnosis of asthma in the DNRP was relatively low at 44%, while the specificity was high at 98%. The predictive value of such a recording did not vary by hospital contact type for asthma (ie, inpatient stay, outpatient clinic contact, emergency department visit). After applying the estimated sensitivity and specificity to an actual cohort study of asthma and risk of skin cancers, we found a more pronounced effect of both direct and inverse associations between asthma and the risk of skin cancers. Only a combined low sensitivity at 20% and specificity at 85% of an asthma diagnosis would be sufficient to nullify observed relative associations. Both values fall well below the estimated sensitivity and specificity of the DNRP records.

Whether the data quality documented in our study is sufficient for registry-based studies depends on the proposed research questions and study design used.¹¹ If the DNRP data is used to assess changes in incidence of asthma over time, sensitivity and specificity must remain stable over time to obtain valid estimates. It is important that the misclassification of asthma is unrelated to information about earlier exposures or future outcomes (ie, nondifferential misclassification), if the DNRP data is used in cohort or

case control studies. We applied the estimates of sensitivity and specificity obtained in this validation study to the results of an actual study, in which we examined the association between asthma and risk of skin cancers (unpublished data). In this study, we found a 13% reduced risk of MM and 13% and 46% increased risk of BCC and SCC, respectively, among asthma patients. Assuming that the asthma misclassification is nondifferential with respect to skin cancer status, we showed that the numeric value of the corrected IRR for MM among asthma patients was reduced, while the numeric values IRR for BCC and SCC among asthma patients were increased without affecting the interpretation; in both cases, the corrected IRR estimate was further away from the value of 1.0 than the misclassified one. This corresponds to the statistical expectation that, given nondifferential misclassification, corrected relative risk estimates are further away from the null than the uncorrected estimates.⁸ Had asthma been misclassified differentially with respect to skin cancer risk, the relative estimates could still be corrected but the direction of the bias would be unpredictable.¹² Moreover, we assumed that the misclassification of skin cancer in the actual study was absent, which may not have been the case, resulting in an unpredictable direction of the bias.

There are several potential explanations for the degree of under-coding of asthma observed in this study, particularly compared with the 0.90 estimate of sensitivity in another Danish study.⁴ One explanation may be that the other Danish study validated the DNRP against medical records,

conventionally considered to be the best choice of a gold standard. Conscript records, used in this study as gold standard, are inferior to the actual medical records in that they themselves may under ascertain asthma diagnoses. That could happen if, for example, another medical condition has already been recorded as the primary reason for exemption. Furthermore, for a disease like asthma, with a considerable clinical spectrum and predominance in childhood, conscripts wanting to serve may not have reported mild cases or early life episodes of asthma. Therefore we may have underestimated the PPV of the DNRP by using conscript records as gold standard. Another explanation may be age, as the previous Danish study validated diagnosis of asthma among hospitalized asthmatic children aged 6–14 years.⁴ As the peak age of onset of asthma is six years,¹³ this diagnosis is the first one to be considered among children with obstructive lung symptoms in paediatric departments. Among adults, other diagnoses may be considered in a patient with first time obstructive lung symptoms. Finally, the asthma diagnosis code in the DNRP has been validated in a population of males who were born in 1977–1983, and therefore for them only inpatient stays with discharge diagnoses of asthma were recorded until 1995. Before 1995, diagnoses made during outpatient clinic visits were not recorded in the DNRP. This, together with the fact that asthma diagnoses made outside hospitals (ie, at general practitioners) may have contributed to the low sensitivity of the asthma diagnoses coding in the DNRP.

In conclusion, we considered a single case of the effect of exposure misclassification on the magnitude of apparent IRR. The low sensitivity of the DNRP ascertained asthma diagnosis is not sufficient to nullify observed relative associations in the actual dataset, and the specificity of DNRP asthma diagnosis is high. While nondifferential misclassification of a dichotomous exposure is expected to produce bias towards the null, provided there is an association, the magnitude of

resulting bias may differ depending on characteristics of exposure and the association under study,¹⁰ and needs to be assessed individually.

Acknowledgment

We are grateful to Professor Timothy L Lash for his advice on this manuscript. The authors report no conflicts of interest in this work.

References

1. Fauci AS, Braunwald E, Kasper DL, et al. editors. Asthma. In: *Harrison's Principles of Internal Medicine*, 17th edition. Columbus, OH: McGraw Hill; 2009;p. 248.
2. Nickelsen TN. Data validity and coverage in the Danish National Health Registry. A literature review. *Ugeskr Laeger*. 2001;164(1):33–37.
3. Sørensen HT, Christensen T, Schlosser HK, Pedersen L. *Use of Medical Databases in Clinical Epidemiology*. 2nd edition. Aarhus, Denmark: SUN-TRYK, Aarhus Universitet; 2009.
4. Moth G, Vedsted P, Schiøtz PO. National registry diagnoses agree with medical records on hospitalized asthmatic children. *Acta Paediatr*. 2007;96(10):1470–1473.
5. Sørensen HT, Christensen T, Schlosser HK, Pedersen L. *Use of Medical Databases in Clinical Epidemiology*. Aarhus, Denmark: Department of Clinical Epidemiology, Aarhus University Hospital; 2008.
6. Andersen TF, Madsen M, Jørgensen J, Møller-Jørgensen 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.
7. Frank L. Epidemiology. When an entire country is a cohort. *Science*. 2000;287(5462):2398–2399.
8. Greenland S. Basic methods for sensitivity analysis of biases. *Int J Epidemiol*. 1996;25(6):1107–1116.
9. Rothman KJ, Greenland S. *Modern Epidemiology*, 2nd ed. Hagerstown, MD: Lippincott-Raven; 1998:347–349.
10. Flegal KM, Brownie C, Haas JD. The effects of exposure misclassification on estimates of relative risk. *Am J Epidemiol*. 1986;123(4):736–751.
11. Sørensen HT, Sabroe S, Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *Int J Epidemiol*. 1996;25(2):435–442.
12. Lash TL. Heuristic thinking and inference from observational epidemiology. *Epidemiology*. 2007;18(1):67–72.
13. Chen W, Mempel M, Schober W, Behrendt H, Ring J. Gender difference, sex hormones, and immediate type hypersensitivity reactions. *Allergy*. 2008;63(11):1418–1427.

Clinical Epidemiology

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventative initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification, systematic

Submit your manuscript here: <http://www.dovepress.com/clinical-epidemiology-journal>

Dovepress

reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.