Considerations re. outcomes

Credit: Maja Hellfritzsch

Outcome / event

Mortality
Suicide attempts
High INR values
Stroke
AMI
Cancer
PCI / CABG
Initiation
Discontinuation
Switching

OUTCOME

Disease

Surgery

Treatment initiation

Biochemical change

Validity?

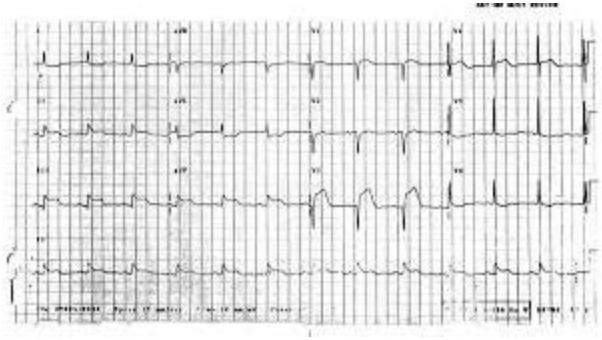
Will this proxy classify those with the outcome as having the outcome? And those without the outcome as not having the outcome?

Is the proxy valid?

Myocardial infarction
= ICD10-code I21

How to test this?

VS.



BMJ Open

Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study

Jens Sundbøll,[1,2] Kasper Adelborg,[1,2] Troels Munch,[1] Trine Frøslev,[1] Henrik Toft Sørensen,[1] Hans Erik Bøtker,[2] Morten Schmidt[1,3]

To cite: Sundbøll J, Adelborg K, Munch T, *et al*. Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study. *BMJ Open* 2016;**6**:e012832. doi:10.1136/bmjopen-2016-012832

► Prepublication history and additional material is available. To view please visit the journal (http://dx.doi.org/10.1136/bmjopen-2016-012832).

Received 26 May 2016
Revised 21 September 2016
Accepted 30 September 2016

ABSTRACT

Objective: The majority of cardiovascular diagnoses in the Danish National Patient Registry (DNPR) remain to be validated despite extensive use in epidemiological research. We therefore examined the positive predictive value (PPV) of cardiovascular diagnoses in the DNPR.

Design: Population-based validation study.

Setting: 1 university hospital and 2 regional hospitals in the Central Denmark Region, 2010–2012.

Participants: For each cardiovascular diagnosis, up to 100 patients from participating hospitals were randomly sampled during the study period using the DNPR.

Main outcome measure: Using medical record review as the reference standard, we examined the PPV for cardiovascular diagnoses in the DNPR, coded according to the International Classification of Diseases, 10th Revision.

Results: A total of 2153 medical records (97% of the total sample) were available for review. The PPVs

Strengths and limitations of this study

- This is the first validation study to include all major cardiovascular diagnoses in the Danish National Patient Registry.
- We sampled patients only from hospitals in the Central Denmark Region. However, our results are most likely generalisable to other parts of the country as the Danish healthcare system is homogeneous in structure and practice.
- We only validated patients diagnosed during 2010–2012 and therefore cannot extrapolate our results to previous periods.

INTRODUCTION

Remarkable improvements have occurred in the prevention and treatment of cardiovascu-

96 of 99 patient with (first)
I21 code had an AMI.

Valid?

	+ Disease	**÷ Disease**
+ Code	True pos.	False pos.
÷ Code	False neg.	True neg.

	+ Disease	**÷ Disease**
+ Code	True pos.	False pos.
÷ Code	False neg.	True neg.

Positive predictive value (PPV):
Likelihood of disease given registration **96 of 99!**

Negative predictive value (NPV):
Likelihood of absence of disease given no registration **?** ≈100%?

Sensitivity (completeness):
Proportion of those with disease having registration **???**

Specificity:
Proportion of those with no disease having no registration **?** ≈100%?

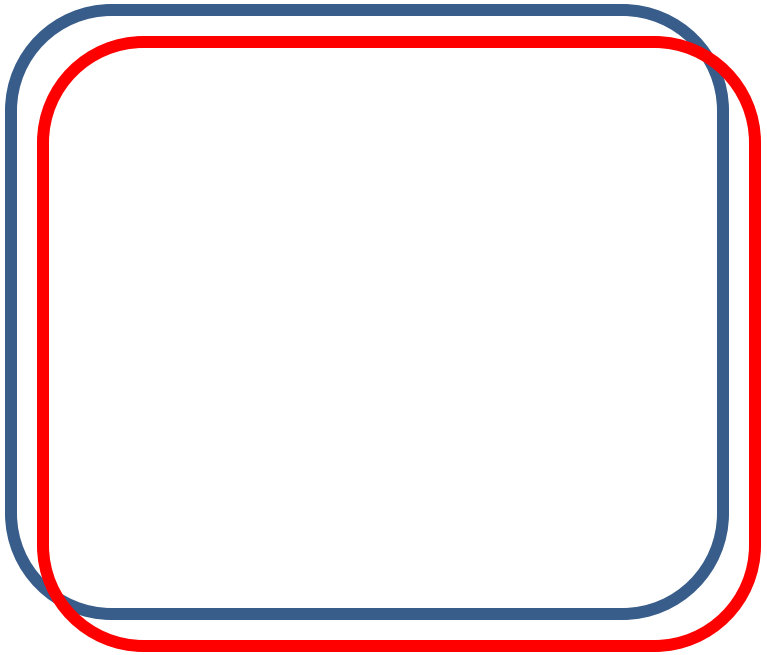
The perfect proxy!

Proxy always represent an outcome
(PPV = 100%)

An outcome will always trigger a proxy
(Sensitivity = 100%)

NOTE: Validation often only adress PPV!

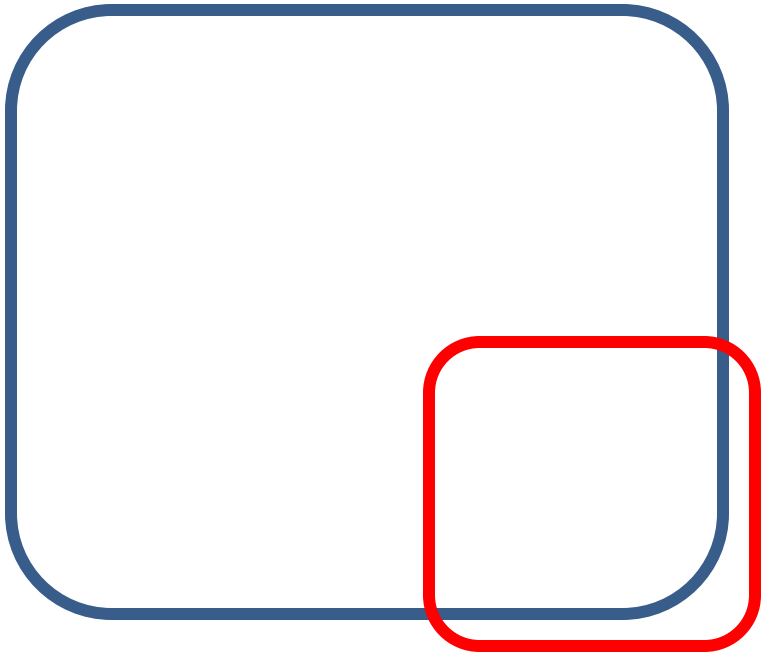
Those with outcome
Those with proxy



High PPV
High sens.

Cancer?

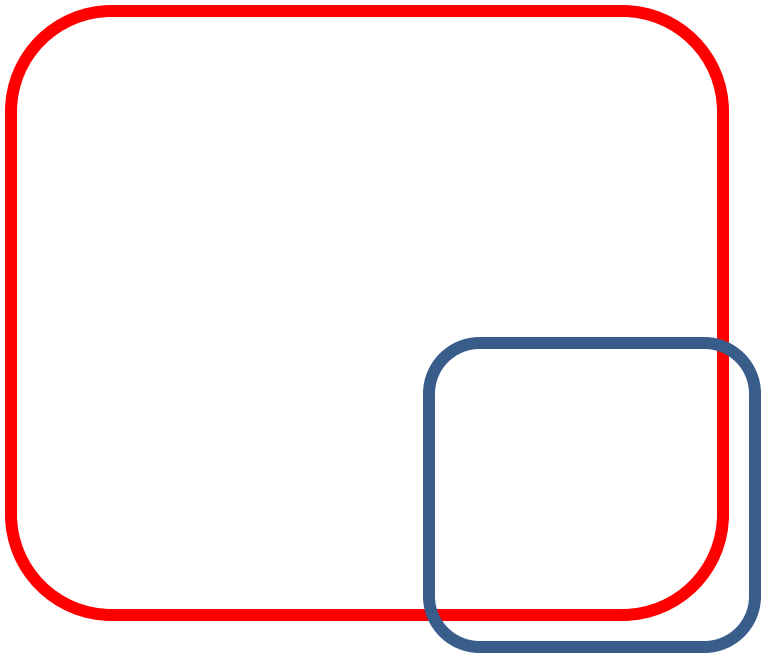
Those with outcome
Those with proxy



High PPV
Low sens.

Obesity diagnosis?

Those with outcome
Those with proxy



Low PPV
High sens.

Gastroscopy as proxy for intestinal bleeding?

ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 323–337

Journal of Clinical Epidemiology

REVIEW ARTICLE

A review of uses of health care utilization databases for epidemiologic research on therapeutics

Sebastian Schneeweiss*, Jerry Avorn

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 1620 Tremont Street (suite 3030), Boston, MA 02120, USA

Accepted 16 October 2004

PPV > Sensitivity

(Most important that the registered outcomes are in fact outcomes!)

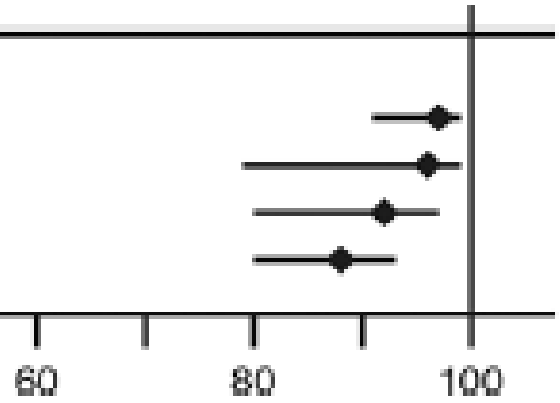
Open Access Research

BMJ Open

Positive predictive value of cardiovascular diagnoses in the Danish National Patient Registry: a validation study

Jens Sundbøll,[1,2] Kasper Adelborg,[1,2] Troels Munch,[1] Trine Frøslev,[1] Henrik Toft Sørensen,[1] Hans Erik Bøtker,[2] Morten Schmidt[1,3]

Myocardial infarction			
First-time myocardial infarction	100	96/99	97 (91-99)
First-time STEMI	23	22/23	96 (79–99)
First-time NSTEMI	39	36/39	92 (80-97)
Recurrent myocardial infarction	100	88/100	88 (80-93)



Suboptimal validity…

Misclassification

What is the height difference between men and women?

Suboptimal validity…

Misclassification of outcome status = information bias

Low PPV →
Those without outcome classified with outcome

Low sensitivity →
Those with outcome classified as not having outcome

As long as validity does not depend on exposure status, misclassification is non-differential and thus biases towards unity (making the groups appear alike)!

How to increase validity?

Algorithms!

Validate!

Stick to codes with high PPV!

Restrict to incident outcomes, primary diagnoses, diagnoses from specialized departments!

Consider sensitivity analyses!

Neuro
-epidemiology

Methods in Neuroepidemiology

Neuroepidemiology 2011;37:120–128
DOI: 10.1159/000331481

Received: March 31, 2011
Accepted: August 2, 2011
Published online: October 7, 2011

Identifying Patients with Myasthenia for Epidemiological Research by Linkage of Automated Registers

Emil Greve Pedersen[a] Jesper Hallas[b] Klaus Hansen[d] Poul Erik Hyldgaard Jensen[e]
David Gaist[a, c]

[a]Department of Neurology, Odense University Hospital, Odense, [b]Institute of Public Health, Clinical Pharmacology Unit, and [c]Institute of Clinical Research, University of Southern Denmark, and [d]Department of Neurology and [e]Neuroimmunology Laboratory, DMSC, Department of Neurology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark

Requiring both diagnosis and prescription yielded PPV of 93%!

Key Words
Myasthenia · Neuromuscular diseases · Neurological disorders · Epidemiology · Research methods

the positive predictive value of the register diagnosis was 92.9% (95% confidence interval, CI, 84.3–97.7), the false-positive rate was low (2.8%), and the sensitivity was acceptable (81.2%; 95% CI 71.2–88.8). ***Conclusions:*** Our data indicate that this novel approach of combining diagnosis register and

Algorithms

Excluding algorithms (increases PPV!)

Multiple requirements to count as outcome

e.g. DVT diagnosis AND later AC treatment

Inclusive algorithms (increases sensitivity!)

Multiple ways of counting as outcome

e.g. diabetes diagnosis OR antidiabetic use

Involve a clinician!

(and beware of pseudo-clinicians!)

Validation?

We defined cases by fulfilment of three criteria: admission with peptic ulcer or gastritis as the main diagnosis to one of the county's hospitals during 1 January 2000 to 31 December 2004; significant bleeding defined by melaena, a subnormal haemoglobin, or the need for transfusions; and a potential bleeding source in the stomach or duodenum identified by endoscopy or surgery.

Hallas et al. BMJ 2006

Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

The Danish National Patient Registry: a review of content, data quality, and research potential

Morten Schmidt[1]
Sigrun Alba Johannesdottir Schmidt[1]
Jakob Lynge Sandegaard[2]
Vera Ehrenstein[1]
Lars Pedersen[1]
Henrik Toft Sørensen[1]

[1]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, [2]Department of Health Documentation, State Serum Institute, Copenhagen, Denmark

Background: The Danish National Patient Registry (DNPR) is one of the world's oldest nationwide hospital registries and is used extensively for research. Many studies have validated algorithms for identifying health events in the DNPR, but the reports are fragmented and no overview exists.

Objectives: To review the content, data quality, and research potential of the DNPR.

Methods: We examined the setting, history, aims, content, and classification systems of the DNPR. We searched PubMed and the *Dantsh Medtcal Journal* to create a bibliography of validation studies. We included also studies that were referenced in retrieved papers or known to us beforehand. Methodological considerations related to DNPR data were reviewed.

Results: During 1977–2012, the DNPR registered 8,085,603 persons, accounting for 7,268,857 inpatient, 5,953,405 outpatient, and 5,097,300 emergency department contacts. The DNPR provides nationwide longitudinal registration of detailed administrative and clinical data. It has recorded information on all patients discharged from Danish nonpsychiatric hospitals since 1977 and on psychiatric inpatients and emergency department and outpatient specialty clinic contacts since 1995. For each patient contact, one primary and optional secondary diagnoses

Table S1 *(Continued)*

ICD codes[a]	Condition	Study period (contact type; diagnosis type)	ICD codes/algorithm[b]	n[c]	PPV; NPV; sensitivity; specificity[d]
I21	Acute myocardial infarction	1996–2009 (IN;* A)	I21	148	PPV =100 (97.5–100)
		1998–2007 (IN/OUT; A)	I21, I22, I23	50	PPV =98.0 (89.5–99.7)
		1993–2003 (IN/OUT/ED; A/B*)	410; I21	1,072	$PPV_{IN/OUT/ED}$ =81.9 (79.5–84.1); $PPV_{IN; A/B}$ =92.4 (90.4–93.9); $PPV_{IN; A}$ =94.4 (92.6–95.7)
		1982–1991 (IN; A/B)	410, 427.24, 427.27, 427.91, 427.97	5,022	PPV_{A} =94.3 (93.6–94.9); PPV_{A+B} =93.4 (92.6–94.0); Se_{A} =62.8 (61.7–64.0); Se_{A+B} =69.5 (68.4–70.6)
		1979–1980 (IN; A/B)	410–414	527	PPV =92.4 (89.8–94.4)
I26	PE	1994–2006 (IN/OUT/ED; A/B)	450.99; I26	353	PPV_{All} =67.4 (62.4–72.1); $PPV_{IN/OUT}$ =82.1 (77.2–86.1); PPV_{ED} =29.6 (22.0–38.5); PPV_{A} =87.0 (81.9–90.9)
	PE during pregnancy and postpartum	1980–2001 (IN;* A*)	450.00–450.99; I26.0–I26.9 + (650–666; O80–84)	22	$PPV_{preg+postpartum}$ =81.8 (59.7–94.8);[e] PPV_{preg} =63.6 (40.7–82.8)[f]
	PE after stroke	2003–2006 (IN; A/B)	I26 (after admission to stroke units and age ≥18 y)	11	PPV =90.9 (62.3–98.4); NPV =97.4 (95.8–98.4); Se =0.0 (0.0–32–4); Sp =100 (99.3–100)
I46	Cardiac arrest	1993–2003 (IN/OUT/ED; A/B*)	427.27; I46	42	$PPV_{IN/OUT/ED}$ =50.0 (35.5–64.5); PPV_{IN} =53.1 (36.5–69.1)
I48	Atrial fibrillation or flutter	1993–2009 (IN/OUT/ED; A/B)	427.93, 427.94; I48	284	PPV_{All} =92.3 (88.6–94.8); $PPV_{IN/OUT}$ =94.0 (90.5–96.3) (independent of diagnosis type and department specialty); PPV_{ED} =64.7 (41.3–82.7)
		1980–2002 (n/a; n/a)	427.93, 427.94; I48	174	PPV =98.9 (95.9–99.7)
		1980–2002 (n/a; n/a)	427.93, 427.94; I48	116	PPV =96.6 (91.5–98.7)
I48.9A	Atrial flutter	1977–1999 (IN/OUT/ED; A/B)	427.94; I48.9A	108	PPV =50.0 (40.7–59.3)
I50	Heart failure	1998–2007 (IN/OUT; A)	I50, I11.0, I13.0, I13.2	50	PPV =100 (92.9–100)

Considerations re validity

What is most important?

To identify all outcomes (high sensitivity)?

To make sure outcomes are correct (high PPV)?

Considerations re validity

What is most important?

To identify all outcomes (high sensitivity)?

To make sure outcomes are correct (high PPV)?

Unless specific considerations:

PPV > Sensitivity