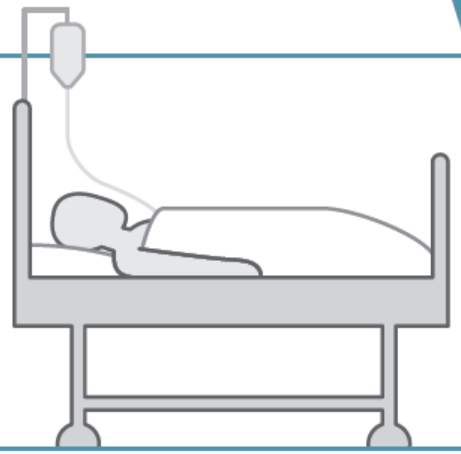
EVERYONE NEEDS VALIDATION

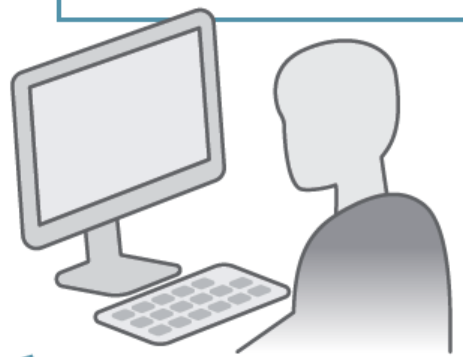
Characteristics of health care system and referral pattern
- Universal coverage
- Insurance-based coverage
- Employment-based coverage
- Access to specialist care



Clinical aspects
- Diagnostic process
- Treatment decisions
- Disease outcome

Generation of RCD/health sector
- Administrative data
- Discharge summaries
- Clinical notes

Type of data/updates
- Structured/coded
- Level of clinical detail
- Database maintenance/ updates/cleaning



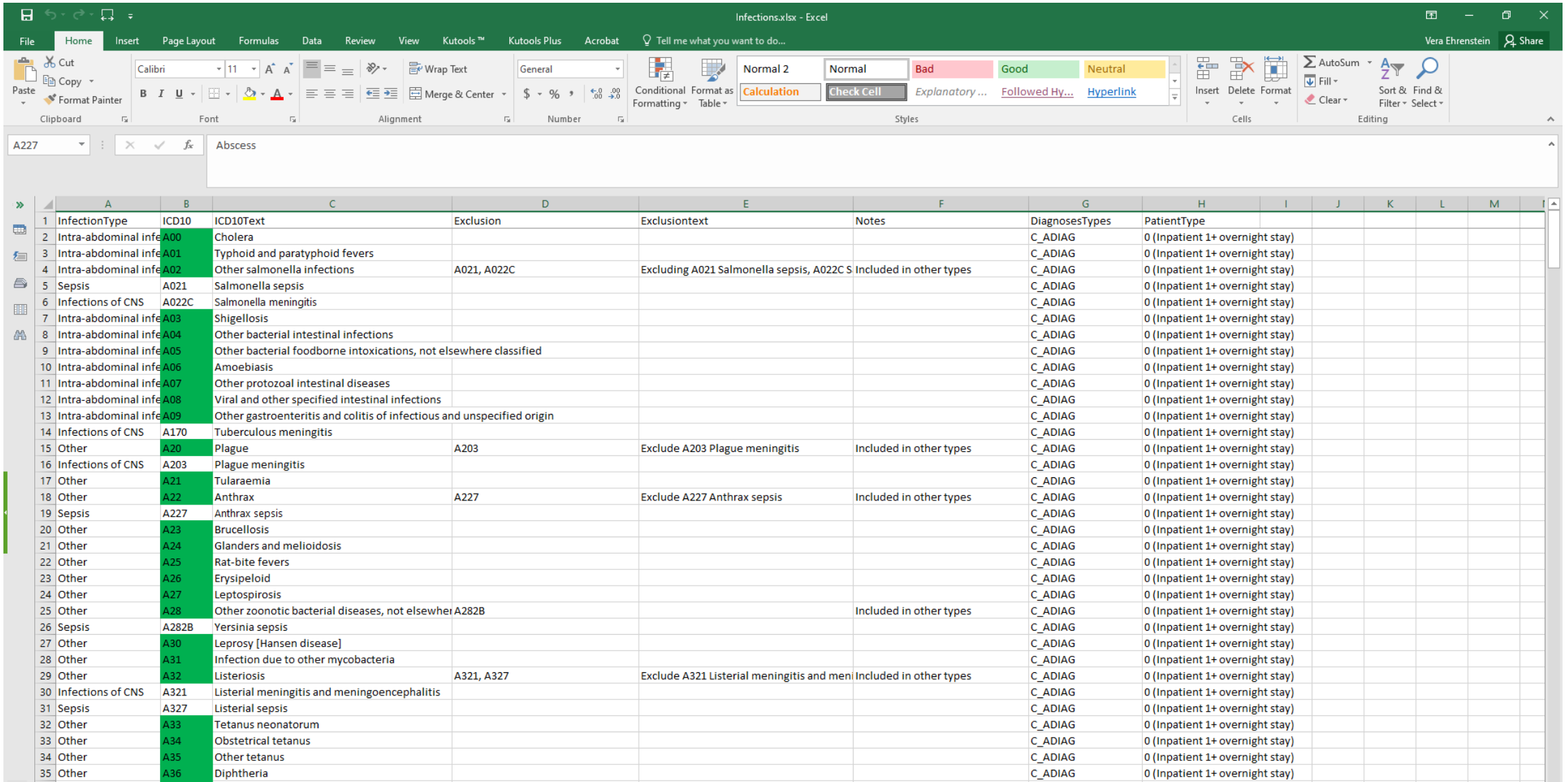
Study conduct
- Study population, design, setting, period
- Choice of RCD algorithms
- Epidemiologic expertise
- Clinical expertise

Evidence generation
- Reporting results
- Interpretation of results

Evidence utilisation
- Meta-analyses
- Guidelines

Patient-level events:
- Symptoms
- Care-seeking thresholds
- Age
- Sex
- Education
- Income

ALGORITHMS - INSTRUMENTS



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	InfectionType	ICD10	ICD10Text	Exclusion	Exclusiontext	Notes	DiagnosesTypes	PatientType					
2	Intra-abdominal infe	A00	Cholera				C_ADIAG	0 (Inpatient 1+ overnight stay)					
3	Intra-abdominal infe	A01	Typhoid and paratyphoid fevers				C_ADIAG	0 (Inpatient 1+ overnight stay)					
4	Intra-abdominal infe	A02	Other salmonella infections	A021, A022C	Excluding A021 Salmonella sepsis, A022C S	Included in other types	C_ADIAG	0 (Inpatient 1+ overnight stay)					
5	Sepsis	A021	Salmonella sepsis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
6	Infections of CNS	A022C	Salmonella meningitis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
7	Intra-abdominal infe	A03	Shigellosis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
8	Intra-abdominal infe	A04	Other bacterial intestinal infections				C_ADIAG	0 (Inpatient 1+ overnight stay)					
9	Intra-abdominal infe	A05	Other bacterial foodborne intoxications, not elsewhere classified				C_ADIAG	0 (Inpatient 1+ overnight stay)					
10	Intra-abdominal infe	A06	Amoebiasis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
11	Intra-abdominal infe	A07	Other protozoal intestinal diseases				C_ADIAG	0 (Inpatient 1+ overnight stay)					
12	Intra-abdominal infe	A08	Viral and other specified intestinal infections				C_ADIAG	0 (Inpatient 1+ overnight stay)					
13	Intra-abdominal infe	A09	Other gastroenteritis and colitis of infectious and unspecified origin				C_ADIAG	0 (Inpatient 1+ overnight stay)					
14	Infections of CNS	A170	Tuberculous meningitis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
15	Other	A20	Plague	A203	Exclude A203 Plague meningitis	Included in other types	C_ADIAG	0 (Inpatient 1+ overnight stay)					
16	Infections of CNS	A203	Plague meningitis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
17	Other	A21	Tularaemia				C_ADIAG	0 (Inpatient 1+ overnight stay)					
18	Other	A22	Anthrax	A227	Exclude A227 Anthrax sepsis	Included in other types	C_ADIAG	0 (Inpatient 1+ overnight stay)					
19	Sepsis	A227	Anthrax sepsis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
20	Other	A23	Brucellosis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
21	Other	A24	Glanders and melioidosis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
22	Other	A25	Rat-bite fevers				C_ADIAG	0 (Inpatient 1+ overnight stay)					
23	Other	A26	Erysipeloid				C_ADIAG	0 (Inpatient 1+ overnight stay)					
24	Other	A27	Leptospirosis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
25	Other	A28	Other zoonotic bacterial diseases, not elsewher	A282B		Included in other types	C_ADIAG	0 (Inpatient 1+ overnight stay)					
26	Sepsis	A282B	Yersinia sepsis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
27	Other	A30	Leprosy [Hansen disease]				C_ADIAG	0 (Inpatient 1+ overnight stay)					
28	Other	A31	Infection due to other mycobacteria				C_ADIAG	0 (Inpatient 1+ overnight stay)					
29	Other	A32	Listeriosis	A321, A327	Exclude A321 Listerial meningitis and meni	Included in other types	C_ADIAG	0 (Inpatient 1+ overnight stay)					
30	Infections of CNS	A321	Listerial meningitis and meningoencephalitis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
31	Sepsis	A327	Listerial sepsis				C_ADIAG	0 (Inpatient 1+ overnight stay)					
32	Other	A33	Tetanus neonatorum				C_ADIAG	0 (Inpatient 1+ overnight stay)					
33	Other	A34	Obstetrical tetanus				C_ADIAG	0 (Inpatient 1+ overnight stay)					
34	Other	A35	Other tetanus				C_ADIAG	0 (Inpatient 1+ overnight stay)					
35	Other	A36	Diphtheria				C_ADIAG	0 (Inpatient 1+ overnight stay)					

ERROR IN EPIDEMIOLOGY

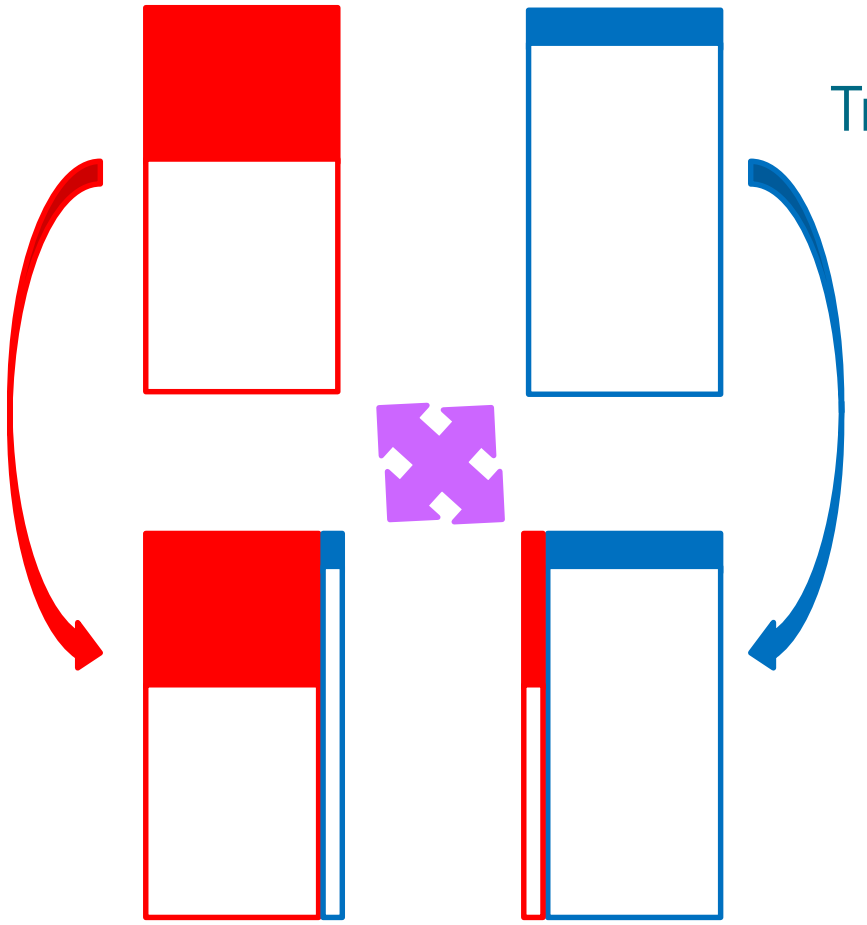
All instruments have errors

- random
- systematic

All epidemiologic measures are instruments

- classifications (eg. BMI to measure obesity)

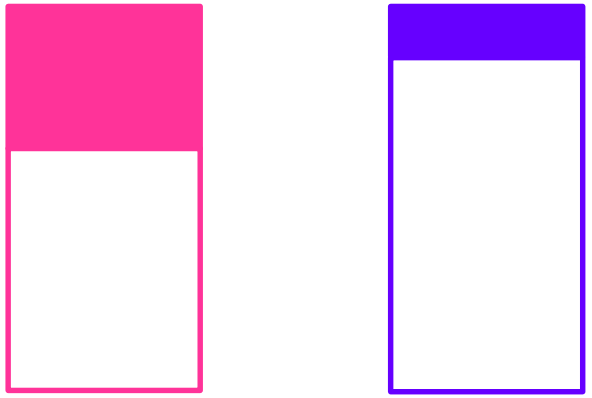
NONDIFFERENTIAL MISCLASSIFICATION OF BINARY EXPOSURE



True RR=4.0

Bias towards the null = dilution

Observed RR=2.85



Slide credit: Tim Lash

MISCLASSIFICATION

- Classifying exposed as unexposed and vice versa
- Classifying cases as noncases and vice versa
- Subgroup/confounder misclassification
- Complex scenarios with > 2 levels, continuous

information bias is a consequence of measurement error

RANDOM ERROR ON THE LEVEL OF A STUDY VARIABLE LEADS TO A **SYSTEMATIC** ERROR ON THE LEVEL OF THE STUDY RESULT

SYSTEMATIC ERROR

Large study size/high precision does not help!

EPIDEMIOLOGY: THE DIAGNOSIS PARADIGM

VALIDATION 2 X 2 TABLE

Gold standard

	Event truly present	Event truly absent	
Algorithm+	a True Positive	b False Positive	a + b
Algorithm-	c False Negative	d True Negative	c + d
	a + c	b + d	

GOLD STANDARD

- Measures status perfectly (gold standard)
- Measures status better than algorithm (alloyed gold / reference standard)
- An independent source of information

- Examples in database validation
  - Medical chart review
  - Surveys
  - Other databases

HYPERCALCEMIA: ICD VS. LAB

	Calcium > 2.6 mmol/L	Calcium <= 2.6 mmol/L
ICD-10 E83.5 Yes	a True Positive	b False Positive
ICD-10 E83.5 No	c False Negative	d True Negative

371. The Validity of Hypercalcemia and Hypocalcemia Diagnoses in Danish Registries

Sussie Antonsen,[1] Christian F Christiansen,[1] Mette Nørgaard,[1] Cathy W Critchlow,[2] Henrik T Sørensen.[1] *[1]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark; [2]Center for observational research, Amgen, Inc., Thousand Oaks, CA, United States.*

Background: Cancer metastasis is one of multiple potential causes of hypocalcemia and hypercalcemia, with clinical manifestations ranging from mild disease to severe hypercalcemia requiring acute treatment. Characterizing the performance of algorithms to identify hypo- and hypercalcemia is important in assuring the validity of pharmacoepidemiological studies using these diagnoses.

Objectives: To evaluate the positive predictive value (PPV) of hypo- and hypercalcemia diagnoses in the Danish National Registry of Patients (DNRP) using the Laboratory Information Systems (LABKA) as reference standard.

Methods: Using data from the DNRP from 1997 through 2006, we identified, among persons living in Northern Den-

Validity of ICD-10 diagnoses of overweight and obesity in Danish hospitals

This article was published in the following Dove Press journal:
Clinical Epidemiology

Sigrid Bjerge Gribsholt[1,2]
Lars Pedersen[1]
Bjørn Richelsen[3]
Reimar Wernich Thomsen[1]

Purpose: Health care databases may be a valuable source for epidemiological research in obesity, if diagnoses are valid. We examined the validity and completeness of International Classification of Diseases, 10th revision [ICD-10] diagnosis coding for overweight/obesity in Danish hospitals.

Total	Confirmed overweight/obesity	Not confirmed overweight/obesity	No BMI recorded	Positive predictive value, (95% CI)
	BMI ≥25 kg/m^2	BMI <25 kg/m^2		
N	n (%)	n (%)	n (%)	%
19,672	15,689 (79.8)	167 (0.8)	3816 (19.4)	79.8 (79.2–80.3)

	DE66 overweight/obesity diagnosis code n (%)	No DE66 overweight/obesity diagnosis code n (%)	Completeness (95% CI)
All	25,619 (10.9)	209,259 (89.1)	10.9 (10.8–11.0)

The data quality of haematological malignancy ICD-10 diagnoses in a population-based Hospital Discharge Registry

M Nørgaard[1,2], M V Skriver[1], H Gregersen[3], G Pedersen[3,4], H C Schønheyder[5] and H T Sørensen[1]

Table 1 Number of patients with a first-time diagnosis of a haematological malignancy in the Hospital Discharge Registry (HDR) in North Jutland County, Denmark, in the Danish Cancer Registry (DCR), and in both registries. Degree of completeness and positive predictive value (PPV) are given as percentages

	Patients registered in					
	Both registries *n* (%)	Only HDR *n* (%)	Only DCR *n* (%)	Total *n*	Degree of completeness % (95% CI)	PPV % (95% CI)
All haematological malignancies	908 (78.3)	167 (14.4)	84 (7.3)	1159	91.5 (89.6–93.1)	84.5 (82.2–86.5)
Acute myeloid leukaemia	73 (62.4)	35 (29.9)	9 (7.7)	117	89.0 (80.4–94.1)	67.6 (58.3–75.7)
Hodgkin's disease	55 (65.5)	22 (26.2)	7 (8.3)	84	88.7 (78.5–94.4)	71.4 (60.5–80.3)
Non-Hodgkin's lymphoma, or chronic lymphocytic leukaemia	523 (76.6)	90 (13.2)	70 (10.3)	683	88.2 (85.3–90.6)	85.3 (82.3–87.9)
Multiple myeloma	130 (76.0)	28 (16.4)	13 (7.6)	171	90.9 (85.1–94.6)	82.3 (75.6–87.4)

Completeness of HDR: 908/(908+84) = 0.915

Eur J Cancer Prev 2005;14:201-6

The data quality of haematological malignancy ICD-10 diagnoses in a population-based Hospital Discharge Registry

M Nørgaard[1,2], M V Skriver[1], H Gregersen[3], G Pedersen[3,4], H C Schønheyder[5] and H T Sørensen[1]

Table 1 Number of patients with a first-time diagnosis of a haematological malignancy in the Hospital Discharge Registry (HDR) in North Jutland County, Denmark, in the Danish Cancer Registry (DCR), and in both registries. Degree of completeness and positive predictive value (PPV) are given as percentages

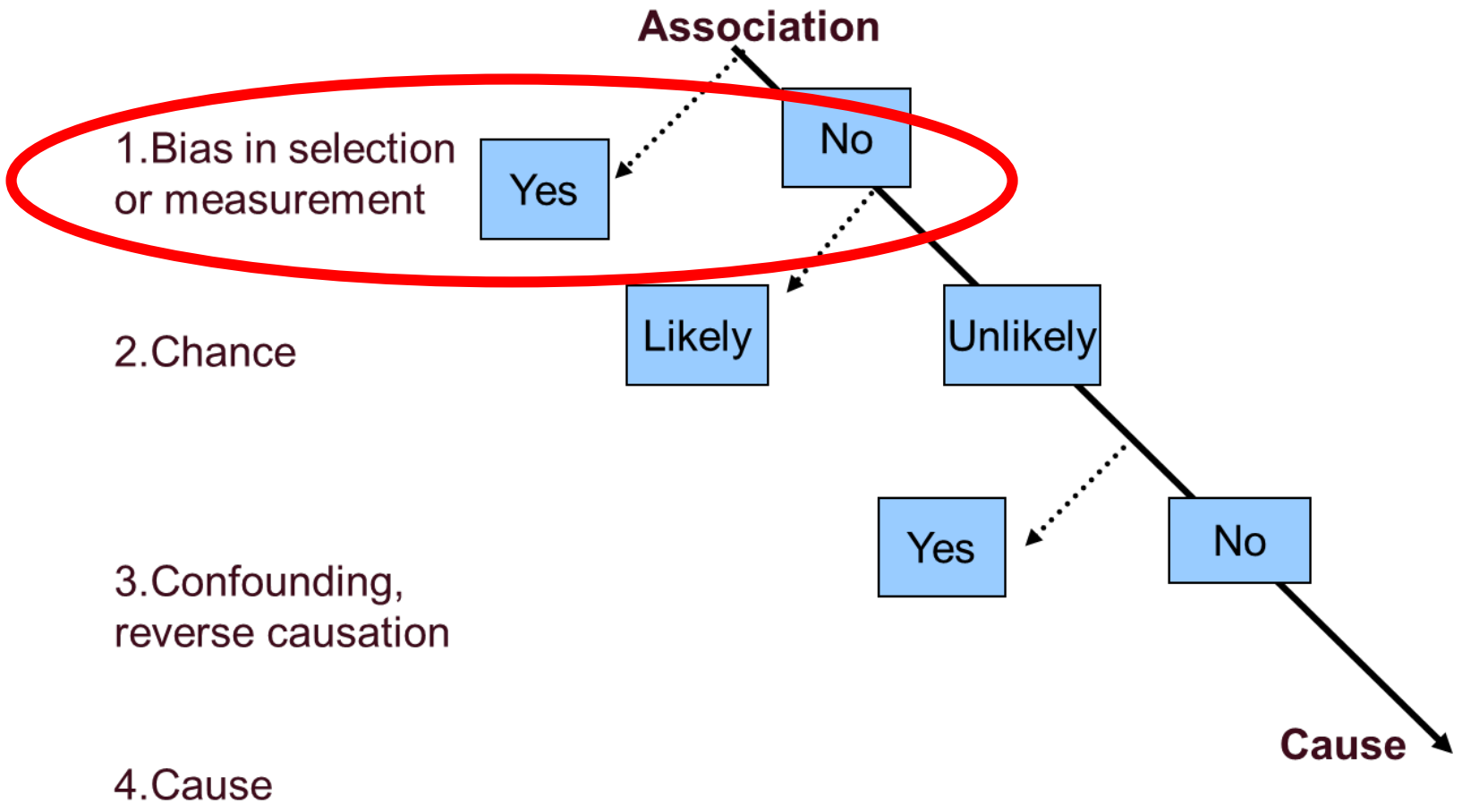
	Patients registered in					
	Both registries *n* (%)	Only HDR *n* (%)	Only DCR *n* (%)	Total *n*	Degree of completeness % (95% CI)	PPV % (95% CI)
All haematological malignancies	908 (78.3)	167 (14.4)	84 (7.3)	1159	91.5 (89.6–93.1)	84.5 (82.2–86.5)
Acute myeloid leukaemia	73 (62.4)	35 (29.9)	9 (7.7)	117	89.0 (80.4–94.1)	67.6 (58.3–75.7)
Hodgkin's disease	55 (65.5)	22 (26.2)	7 (8.3)	84	88.7 (78.5–94.4)	71.4 (60.5–80.3)
Non-Hodgkin's lymphoma, or chronic lymphocytic leukaemia	523 (76.6)	90 (13.2)	70 (10.3)	683	88.2 (85.3–90.6)	85.3 (82.3–87.9)
Multiple myeloma	130 (76.0)	28 (16.4)	13 (7.6)	171	90.9 (85.1–94.6)	82.3 (75.6–87.4)

PPV of HDR: 908/(908+167) = 0.844

Eur J Cancer Prev 2005;14:201-6

WE RELY ON ALGORITHMS TO DEFINE

- Eligibility criteria
- Exposures
- Outcomes
- Covariates
- Subgroups



Fletcher RH, Fletcher SW. *Clinical epidemiology : the essentials*. 4th ed. Philadelphia: Lippincott Williams & Wilkins; 2005.

VALIDATION STUDIES = QUANTIFICATION OF MEASUREMENT ERROR

Error is bad

Error is (almost always) unavoidable

How bad is it?

How much error is acceptable?

Can/should it be corrected?

What is the impact on the interpretation?

International Journal of Epidemiology

Vol. 25, No. 2
Printed in Great Britain

A Framework for Evaluation of Secondary Data Sources for Epidemiological Research

HENRIK TOFT SØRENSEN,* SVEND SABROE AND JØRN OLSEN†**

Sørensen H T (Department of Internal Medicine V, Aarhus University Hospital DK-8000 Aarhus C, Denmark) Sabroe S and Olsen J. A framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology* 1996; **25:** 435–442.

Background. As part of the development in information technology, increasing amounts of health care data are available for epidemiological research.

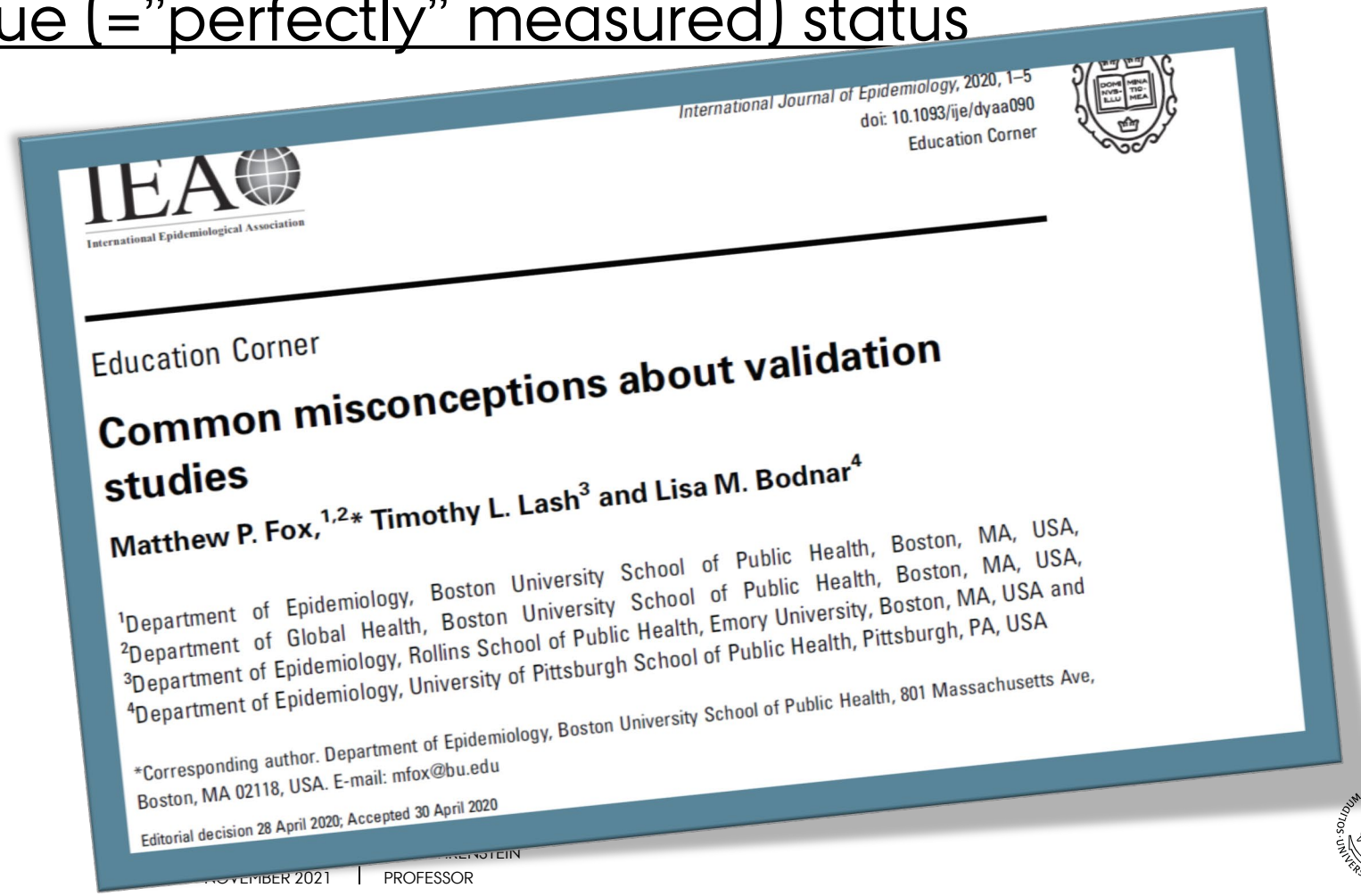
Methods. In this review, we discuss the following factors affecting the value of secondary data in research: 1) completeness of registration of individuals, 2) the accuracy and degree of completeness of the registered data, 3) the size of the data source, 4) the registration period, 5) data accessibility, availability and cost, 6) data format, and 7) possibilities of linkage with other data sources (record linkage).

Results and Conclusion. The importance of these issues depends on the use of the data and on the problems they have to address. If the evaluation is satisfactory with respect to the above-mentioned factors relevant to the particular study, the data source could be a very cost-effective way of solving the research problem.

Keywords: evaluation, validity, accuracy, information systems, data sources, records

VALIDATION STARTING POINT

- Design 1: sample on potential (=”imperfectly” measured) status
- Design 2: sample on true (=”perfectly” measured) status
- Design 3: sample both



International Journal of Epidemiology, 2020, 1–5
doi: 10.1093/ije/dyaa090
Education Corner

IEA
International Epidemiological Association

Education Corner

Common misconceptions about validation studies

Matthew P. Fox,[1,2]* Timothy L. Lash[3] and Lisa M. Bodnar[4]

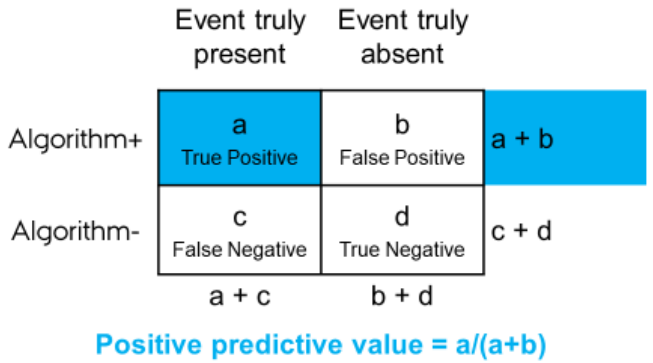
[1]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA,
[2]Department of Global Health, Boston University School of Public Health, Boston, MA, USA,
[3]Department of Epidemiology, Rollins School of Public Health, Emory University, Boston, MA, USA and
[4]Department of Epidemiology, University of Pittsburgh School of Public Health, Pittsburgh, PA, USA

*Corresponding author. Department of Epidemiology, Boston University School of Public Health, 801 Massachusetts Ave, Boston, MA 02118, USA. E-mail: mfox@bu.edu

Editorial decision 28 April 2020; Accepted 30 April 2020

VALIDATION DESIGN 1: POTENTIAL STATUS

- Define candidate event-finding algorithm (e.g., ICD-10)
- Identify a 'gold standard' source (e.g., charts)
- Define criteria for events and non-events
- Determine study size/precision of estimates
- Sample potential cases using algorithm (e.g., ICD-10 codes)
- Conduct validation (standard CRF)
- Estimate **positive/negative predictive value**



Positive predictive values of cardiovascular diagnoses in the Danish National Patient Registry.

Study population

We used the DNPR to randomly sample inpatient and outpatient hospital diagnoses from the Central Denmark Region between 1 January 2010 and 31 December 2012.

Our study population consisted of patients discharged with a primary or secondary first-time diagnosis from departments of cardiology, internal medicine, acute medicine and neurology in the three hospitals. For myocardial infarction, heart failure and venous thromboembolism, we also validated recurrent events.

Medical record review

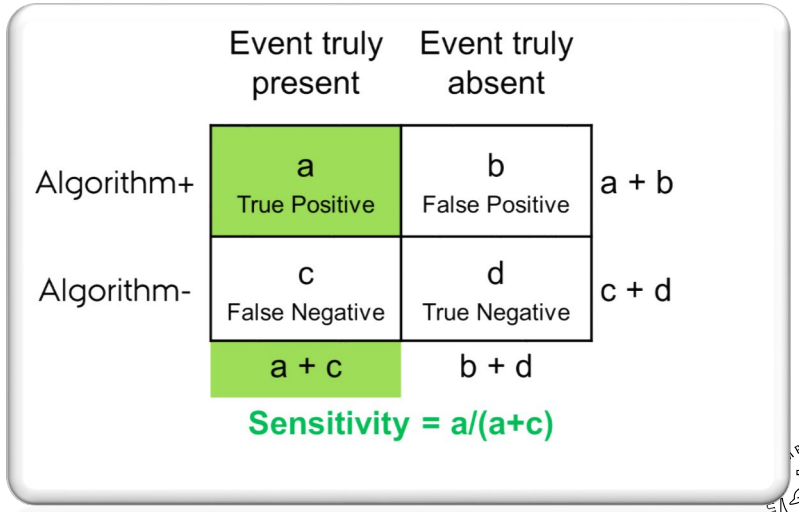
Medical record review was used as the reference standard. We did not have access to ECGs or other paraclinical recordings that supported the clinician's decision. However, descriptions of such recordings were available in the medical records and included in the review process. Three physicians (JS, KA and TM) reviewed the medical records and judged whether they confirmed the cardiovascular diagnosis coded in the DNPR. If the diagnosis was not described in the discharge summary or if the discharge summary was not available, the full medical record was reviewed to examine whether the diagnosis code could be confirmed.

Disease	Sample	Ratio	PPV, % (95% CI)
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)
Stent thrombosis	24	22/24	92 (74-98)
Angina pectoris			
Stable angina pectoris	100	89/96	93 (86-96)
Unstable angina pectoris	100	84/96	88 (79-93)
Heart failure			
First-time admission	100	72/95	76 (66-83)
Readmission	100	73/96	76 (67-83)
Cardiomyopathy			
Cardiomyopathy overall	89	80/89	90 (82-95)
Dilated cardiomyopathy	20	15/20	75 (53-89)
Hypertrophic cardiomyopathy	20	18/20	90 (70-97)
Restrictive cardiomyopathy	9	7/9	78 (45-94)
Arrhythmogenic right ventricular cardiomyopathy	20	20/20	100 (84-100)
Takotsubo cardiomyopathy	20	20/20	100 (84-100)
Hypertension			
Arterial hypertension	100	89/97	92 (85-96)
Pulmonary hypertension	100	87/100	87 (79-92)
Arrhythmia			
Atrial fibrillation or flutter	100	92/97	95 (89-98)
Bradycardia	100	87/100	87 (79-92)
Ventricular tachycardia or fibrillation	100	77/96	80 (71-87)
Cardiac arrest	100	94/100	94 (88-97)
Valvular heart disease			
Mitral regurgitation or stenosis	50	47/49	96 (86-99)
Aortic regurgitation or stenosis	50	49/50	98 (90-100)
Inflammation/Infection			
Endocarditis	100	79/96	82 (73-89)
Myocarditis	73	42/66	64 (52-74)
Pericarditis	100	90/98	92 (85-96)
Aortic diseases			
Aortic dissection	50	46/50	92 (81-97)
Aortic aneurysm or dilatation	50	50/50	100 (93-100)
Venous thromboembolism			
First-time venous thromboembolism	100	87/99	88 (80-93)
First-time deep venous thrombosis	51	43/50	86 (74-93)
First-time pulmonary embolism	49	44/49	90 (78-96)
Recurrent venous thromboembolism	100	67/93	72 (62-80)
Recurrent deep venous thrombosis	41	29/39	74 (59-85)
Recurrent pulmonary embolism	59	38/54	70 (57-81)
Other			
Arterial claudication	100	88/97	91 (83-95)
Hypercholesterolemia	100	90/94	96 (90-98)
Cardiac tumors	26	22/26	85 (66-94)

60 80 100

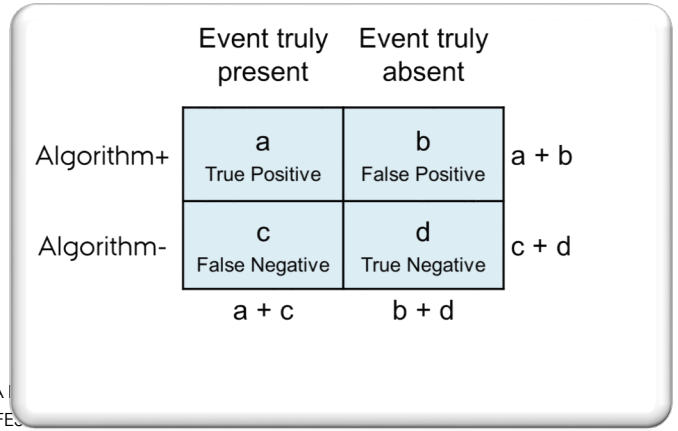
VALIDATION DESIGN 2: TRUE STATUS

- Define candidate event-finding algorithm (e.g., ICD-10)
- Identify an independent sample of true events and non-events ('a gold standard')
- Define criteria for events and non-events
- Search the database for algorithm components among the members of the independent sample
- Conduct validatio (CRF)
- Compute algorithm **sensitivity/specificity**
  - Any code
  - Each code



VALIDATION DESIGN 3: STATUS-INDEPENDENT

- Define candidate event-finding algorithm (e.g., ICD-10)
- Define a gold standard
- Define criteria for events and non-events
- Determine algorithm-based event/non-event status
- Compute **sensitivity, specificity, PPV, NPV** of the event-finding algorithm
  - Any component
  - Each component



Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Validity of the Danish National Registry of Patients for chemotherapy reporting among colorectal cancer patients is high

This article was published in the following Dove Press journal:
Clinical Epidemiology
29 August 2013
Number of times this article has been viewed

Jennifer L Lund[1]
Trine Frøslev[1]
Thomas Deleuran[1,2]
Rune Erichsen[1]
Tove Nilsson[1]
Annette Nørkær Pedersen[3]
Morten Høyer[4]

[1]Department of Clinical Epidemiology, [2]Department of Medicine V (Hepatology and Gastroenterology), [3]Hospital Pharmacy, [4]Department of Oncology, Aarhus University Hospital, Aarhus, Denmark

Background: The Danish National Registry of Patients (DNRP) is a potentially valuable resource for monitoring national trends in the use of chemotherapy and evaluating the benefits and harms of alternative treatments among colorectal cancer (CRC) patients in Denmark. However, the validity of chemotherapy reporting in the DNRP is unknown. In this study, we evaluated the validity of the DNRP for identifying the receipt of chemotherapy and specific treatments, and the timing and number of treatments among CRC patients, using medical records and pharmacy data as the reference standard.

Methods: We selected a random sample of CRC patients with lymph node involvement who were diagnosed at Aarhus University Hospital (n = 25) or Aalborg University Hospital (n = 25) from 2009 to 2010. Administration dates, specific treatments, and number of treatment courses were abstracted for the 180 days post diagnosis from the DNRP, medical records, and pharmacy production databases. The prevalence of chemotherapy, timing of first administration, and number of courses were described. DNRP data were compared with the reference standard for each hospital, and the kappa, sensitivity, specificity, positive and negative predictive values, and 95% confidence intervals were calculated for the receipt of any chemotherapy and specific treatments.

Results: The prevalence of chemotherapy was 72% and 68% among CRC patients treated in Aarhus and Aalborg, respectively, with >90% of patients without distant metastasis receiving treatment within 90 days from diagnosis. Patients received on average 4.6 and 4.7 treatment courses in Aarhus and Aalborg, respectively. Kappa, sensitivity, and specificity of chemotherapy reporting in the DNRP was high (≥0.88), but the sensitivity of individual chemotherapies varied by hospital.

Conclusion: The validity of chemotherapy reporting in the DNRP was high, although some variation by hospital exists. The DNRP represents a population-based nationwide resource that can be used to provide timely and accurate evaluations of chemotherapy use among CRC patients in Denmark.

Keywords: chemotherapy, colorectal cancer, administrative data

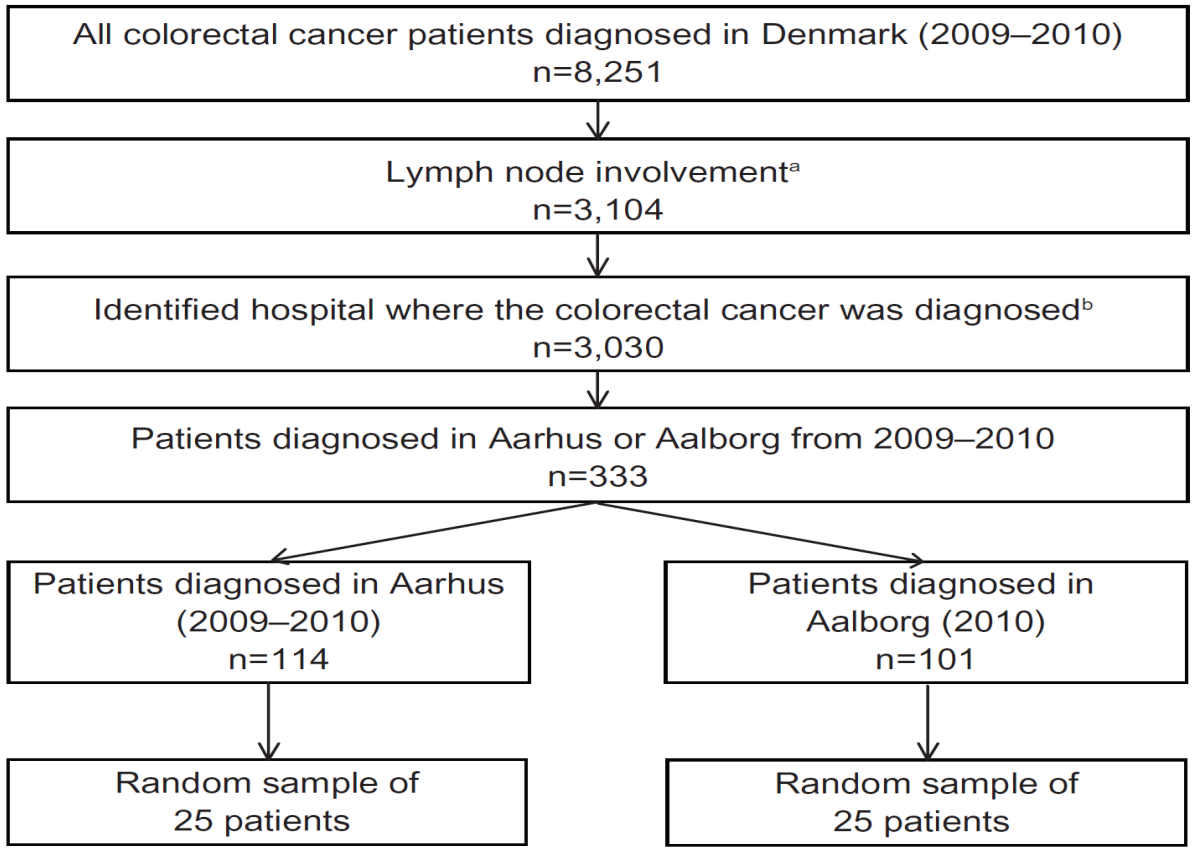


Figure 1 Study flow chart of patients selected for inclusion in the validation study.
Notes: [a]Regional lymph node involvement was identified using the tumor, node, metastasis (TNM) staging system (N ≥ N1); [b]using an administrative algorithm requiring a hospital diagnosis of colorectal cancer (International Classification of Diseases (ICD)-10 codes: C18-20) ± 14 days from the diagnosis date, as reported by the Danish Cancer Registry.

Table 2 Measures of agreement and validity comparing reporting of chemotherapy in the Danish National Registry of Patients with the reference standard by hospital

Treatment	Source of reporting				Sensitivity 95% CI	Specificity 95% CI	PPV 95% CI	NPV 95% CI	Kappa 95% CI
	Ref Y DNRP Y	Ref N DNRP N	Ref Y DNRP N	Ref N DNRP Y					
Aarhus									
Chemotherapy	17	7	1	0	0.94 (0.84–1.00)	1.00	1.00	0.87 (0.65–1.00)	0.90 (0.72– 1.00)
5-FU[a]	0	23	2	0	0.00	1.00	–	0.92 (0.81–1.00)	–
Oxaliplatin[b]	3	13	9	0	0.25 (0.05–0.50)	1.00	1.00	0.59 (0.39–0.80)	0.26 (0.00–0.52)
Bevacizumab[c]	3	21	1	0	0.75 (0.33–1.00)	1.00	1.00	0.95 (0.87–1.00)	0.83 (0.52–1.00)
Aalborg									
Chemotherapy	17	7	0	1	1.00	0.88 (0.65–1.00)	0.94 (0.84–1.00)	1.00	0.90 (0.72–1.00)
5-FU[a]	8	16	0	1	1.00	0.94 (0.83–1.00)	0.89 (0.68–1.00)	1.00	0.91 (0.74–1.00)
Oxaliplatin[b]	3	22	0	0	1.00	1.00	1.00	1.00	1.00
Bevacizumab[c]	6	19	0	0	1.00	1.00	1.00	1.00	1.00

Notes: [a]Includes administration of 5-FU or capecitabine without oxaliplatin or bevacizumab; [b]includes administration of oxaliplatin in combination with 5-FU (FOLFOX) or capecitabine (XELOX), but without bevacizumab; [c]includes administration of bevacizumab in combination with FOLFOX, XELOX, irinotecan and 5-FU (FOLFIRI), or 5-FU.

Abbreviations: CI, confidence interval; Ref, reference standard; Y, yes; N, no; DNRP, Danish National Registry of Patients; PPV, positive predictive value; NPV, negative predictive value; 5-FU, 5-fluorouracil.

PRIORITIZING VALIDITY MEASURES

ORIGINAL RESEARCH

Validity of ICD-10 diagnoses of overweight and obesity in Danish hospitals

This article was published in the following Dove Press journal:
Clinical Epidemiology

Sigrid Bjerge Gribsholt[1,2]
Lars Pedersen[1]
Bjørn Richelsen[3]
Reimar Wernich Thomsen[1]

Purpose: Health care databases may be a valuable source for epidemiological research in obesity, if diagnoses are valid. We examined the validity and completeness of International Classification of Diseases, 10th revision [ICD-10] diagnosis coding for overweight/obesity in Danish hospitals.

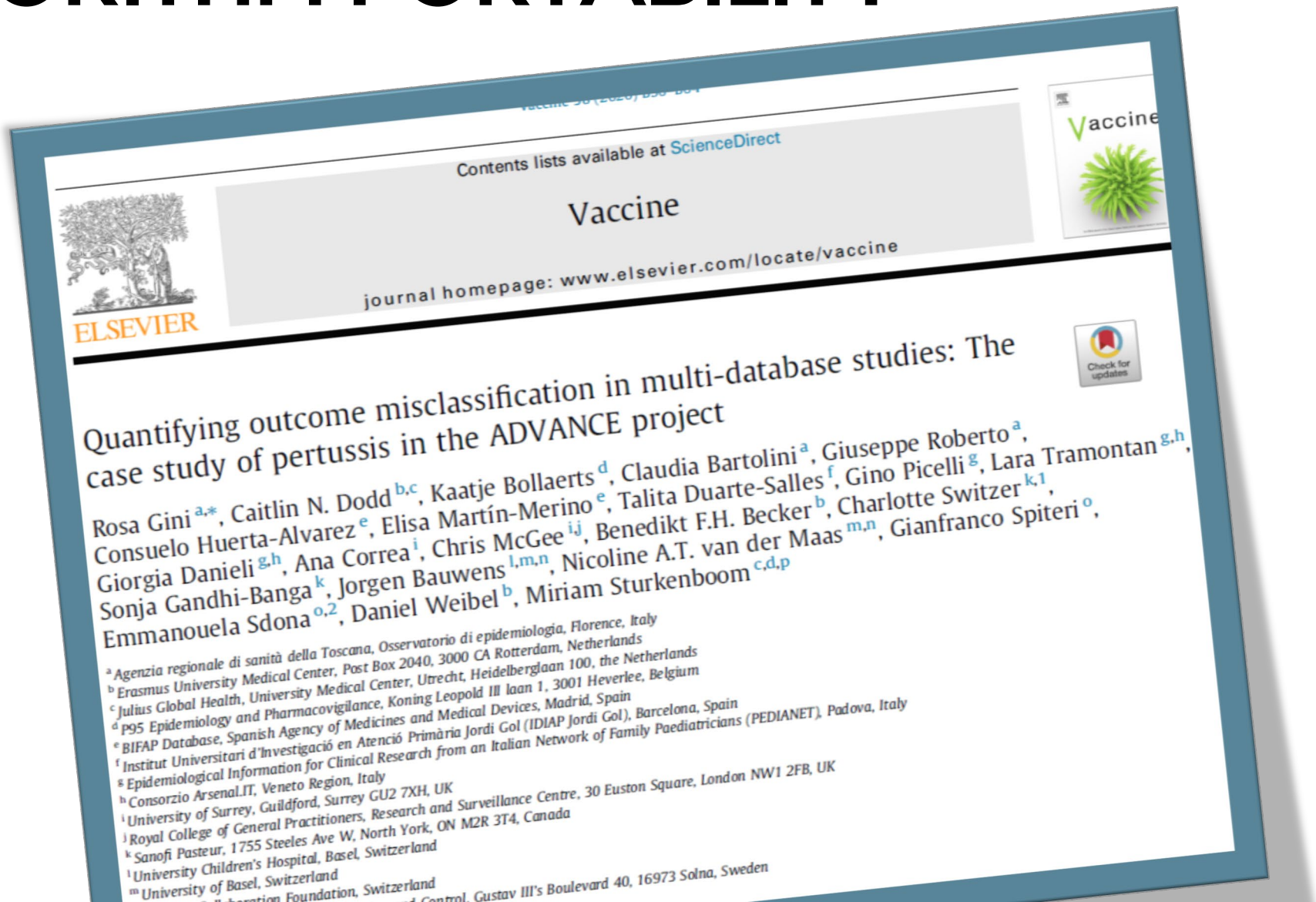
Total	Confirmed overweight/obesity	Not confirmed overweight/obesity	No BMI recorded	Positive predictive value, (95% CI)
	BMI ≥25 kg/m^2	BMI <25 kg/m^2		
N	n (%)	n (%)	n (%)	%
19,672	15,689 (79.8)	167 (0.8)	3816 (19.4)	79.8 (79.2–80.3)

	DE66 overweight/ obesity diagnosis code n (%)	No DE66 overweight/ obesity diagnosis code n (%)	Completeness (95% CI)
All	25,619 (10.9)	209,259 (89.1)	10.9 (10.8–11.0)

PRIORITIZING VALIDITY MEASURES

- **Eligibility** criteria – may prioritize specificity/indication
- **Exposure**
  - Strive to make errors non-differential (e.g., blinding)
- **Outcome**
  - May prioritize specificity for relative risks
  - May prioritize sensitivity for absolute risks
- **Confounders/covariates**
    - May prioritize sensitivity (confounding excluded is confounding uncontrolled)

ALGORITHM PORTABILITY



Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

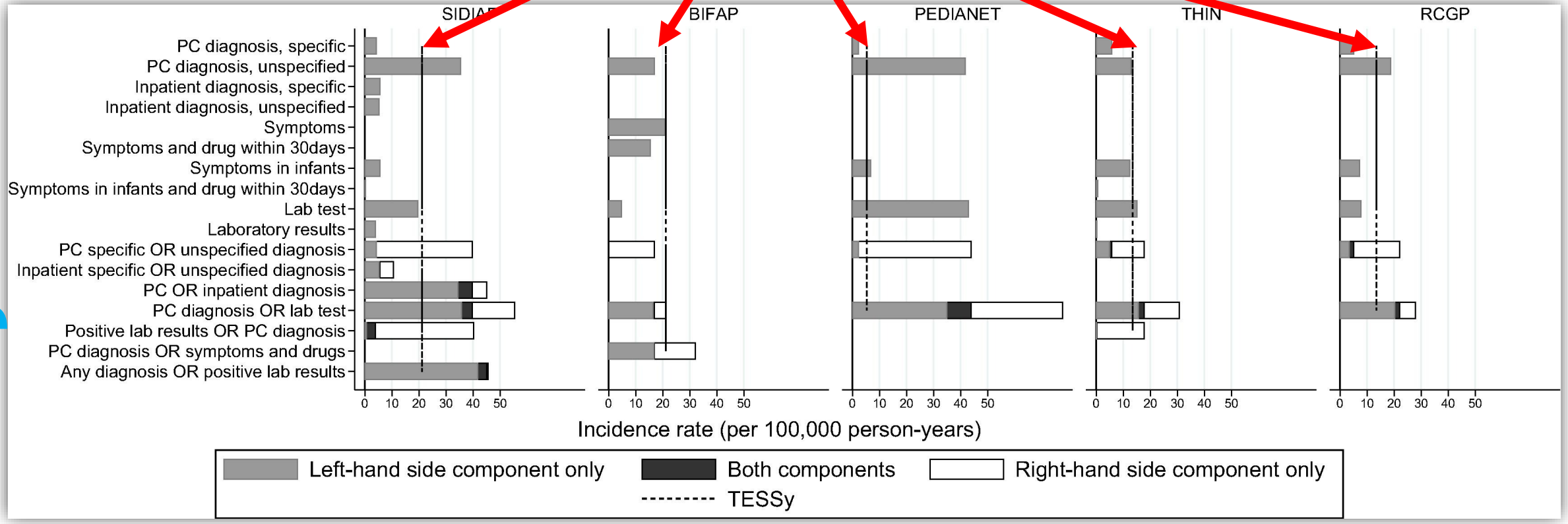
Quantifying outcome misclassification in multi-database studies: The case study of pertussis in the ADVANCE project

Rosa Gini [a,*], Caitlin N. Dodd [b,c], Kaatje Bollaerts [d], Claudia Bartolini [a], Giuseppe Roberto [a], Consuelo Huerta-Alvarez [e], Elisa Martín-Merino [e], Talita Duarte-Salles [f], Gino Picelli [g], Lara Tramontan [g,h], Giorgia Danieli [g,h], Ana Correa [i], Chris McGee [i,j], Benedikt F.H. Becker [b], Charlotte Switzer [k,1], Sonja Gandhi-Banga [k], Jorgen Bauwens [l,m,n], Nicoline A.T. van der Maas [m,n], Gianfranco Spiteri [o], Emmanouela Sdona [o,2], Daniel Weibel [b], Miriam Sturkenboom [c,d,p]

[a] Agenzia regionale di sanità della Toscana, Osservatorio di epidemiologia, Florence, Italy
[b] Erasmus University Medical Center, Post Box 2040, 3000 CA Rotterdam, Netherlands
[c] Julius Global Health, University Medical Center, Utrecht, Heidelberglaan 100, the Netherlands
[d] P95 Epidemiology and Pharmacovigilance, Koning Leopold III laan 1, 3001 Heverlee, Belgium
[e] BIFAP Database, Spanish Agency of Medicines and Medical Devices, Madrid, Spain
[f] Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Barcelona, Spain
[g] Epidemiological Information for Clinical Research from an Italian Network of Family Paediatricians (PEDIANET), Padova, Italy
[h] Consorzio Arsenal.IT, Veneto Region, Italy
[i] University of Surrey, Guildford, Surrey GU2 7XH, UK
[j] Royal College of General Practitioners, Research and Surveillance Centre, 30 Euston Square, London NW1 2FB, UK
[k] Sanofi Pasteur, 1755 Steeles Ave W, North York, ON M2R 3T4, Canada
[l] University Children's Hospital, Basel, Switzerland
[m] University of Basel, Switzerland
[n] Vaccination Foundation, Switzerland
[o] European Centre for Disease Prevention and Control, Gustav III's Boulevard 40, 16973 Solna, Sweden

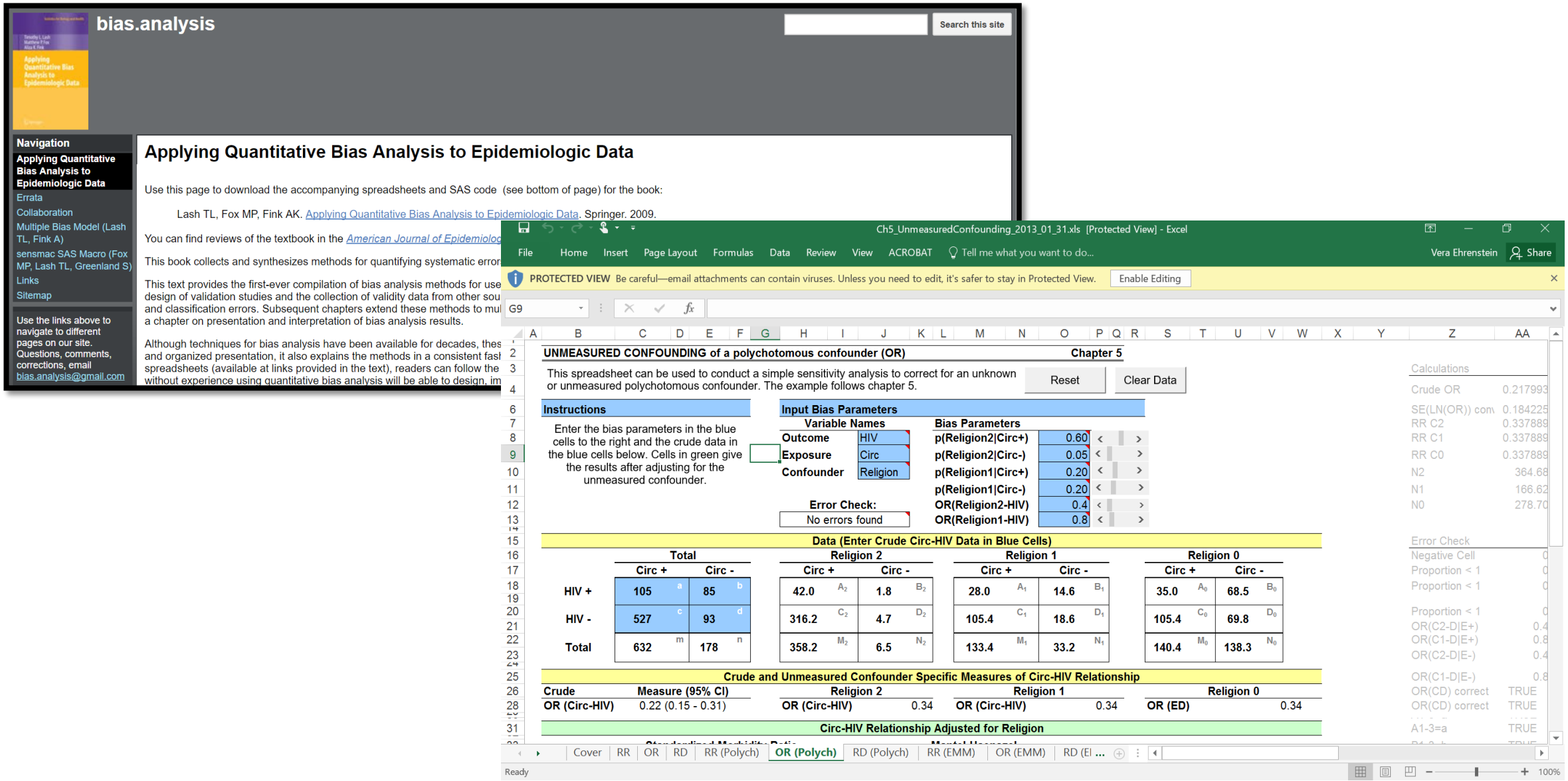
Gold standard

Algorithms



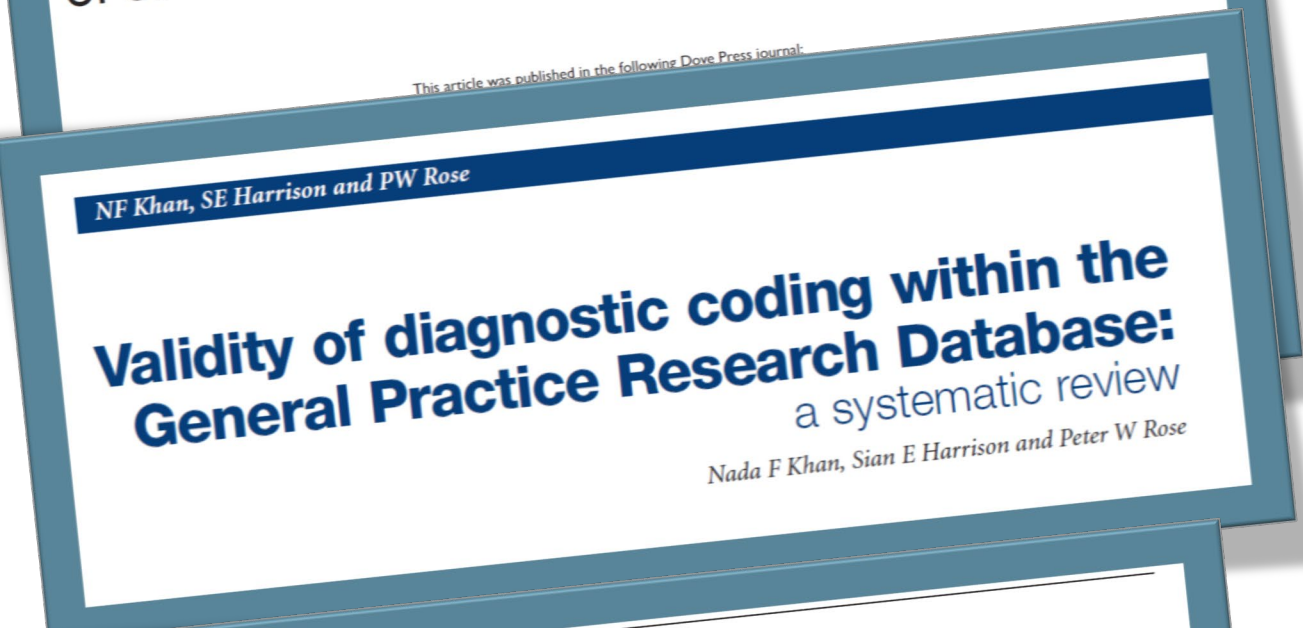
OVERARCHING PRINCIPLES

- Accept the reality that errors are inevitable
- When prioritizing validity metrics, consider consequences of the error to interpretation/clinical practice
- Try to quantify error impact – bias analysis



https://sites.google.com/site/biasanalysis/


Clinical Epidemiology
Dovepress
open access to scientific and medical research
REVIEW
Open Access Full Text Article
The Danish National Patient Registry: a review of content, data quality, and research potential


NF Khan, SE Harrison and PW Rose
Validity of diagnostic coding within the General Practice Research Database:
a systematic review
Nada F Khan, Sian E Harrison and Peter W Rose


WILEY
DOI: 10.1002/pds.4694
COMMENTARY
Pharmacoepidemiology and Drug Safety's special issue on validation studies
Danielle S. Chun | Jennifer L. Lund | Til Stürmer
Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, North Carolina, USA
Correspondence
D. S. Chun, Epidemiology, University of North Carolina at Chapel Hill Gillings School of Global Public Health, Chapel Hill, NC, USA.


Clinical Epidemiology
Dovepress
open access to scientific and medical research
EDITORIAL
Open Access Full Text Article
Helping everyone do better: a call for validation studies of routinely recorded health data
This article was published in the following Dove Press journal:
Clinical Epidemiology
12 April 2016
Number of times this article has been viewed
Vera Ehrenstein[1]
Irene Petersen[1,2]
Liam Smeeth[3]
Susan S Jick[4]
Eric I Benchimol[5,6]
Jonas F Ludvigsson[7,8]
There has been a surge of availability and use for research of routinely collected electronic health data, such as electronic health records, health administrative data, and disease registries. Symptomatic of this surge, in 2012, Pharmacoepidemiology and Drug Safety (PDS) published a supplemental issue containing several reviews of validated methods for identifying health outcomes using routine health data,[1] focusing on databases feeding the US Mini-Sentinel Program.[2] In one of the review


EPIDEMIOLOGY
Articles & Issues
Collections
Multimedia
For Authors
Journal Info
Next Article >
EPIDEMIOLOGY Announces the "Validation Study" Submission Category
Lash, Timothy L.; Olshan, Andrew F.
Epidemiology: September 2016 - Volume 27 - Issue 5 - p 613–614
doi: 10.1097/EDE.0000000000000532
Commentary