Appendix

MEDLINE (via Pubmed) search strategy

1. (health administrative) OR (administrative data) OR (administrative database) OR (claim administrative) OR (International Classification of Diseases) OR "International Classification of Diseases"[Mesh] OR ICD-9-CM OR ICD-10 OR "Database Management Systems"[Mesh] OR "Medical Records Systems, Computerized"[Mesh] OR "CPT" OR "Current procedural terminology"[Mesh] OR (computerized medical records systems) OR (electronic healthcare record) OR (computerized medical record) OR (electronic medical record) OR (automated medical record) OR (electronic patient record) OR CPRD OR GPRD OR Optum OR PHARMO OR HealthCore OR Danish registries

2. (factual databases) OR (geographic information systems) OR (national practitioner data bank) OR (insurance database)

3. #1 OR #2

4. sensitivity or "Sensitivity and Specificity"[Mesh]

5. specificity[Title/Abstract]

6. (positive predictive value) OR (negative predictive value) OR (likelihood ratio) OR (receiver operating characteristic) OR kappa

7. ((case or cases) AND (verificat* OR valid* OR identif* OR definition* OR define* OR evaluat*))

8. Algorithm OR "Algorithm"[Mesh]

9. #4 OR #5 OR #6 OR #7 OR #8

10. emphysema

11. (chronic* bronchitis*)

12. (obstruct* (pulmonary or lung* or airway* or airflow* or bronch* or respirat*))

13. (chronic obstructive pulmonary disease)[Text Word]

14. COPD

15. CAPD

16. "pulmonary disease, chronic obstructive"[MeSH Terms]

17. #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16

18. #3 AND #9 AND #17

EMBASE search strategy (via embase.com)

1. health NEAR/3 administrative OR administrative NEAR/3 data OR administrative NEAR/3 database OR claim NEAR/3 administrative OR (International Classification of Diseases) OR 'International Classification of Diseases'/exp OR ICD-9-CM OR ICD-10 OR 'Database Management Systems'/exp OR 'Medical Records Systems, Computerized'/exp OR 'CPT' OR 'Current procedural terminology'/exp OR (computerized medical records systems) OR (electronic healthcare record) OR (computerized medical record) OR (electronic medical record) OR (automated medical record) OR (electronic patient record) OR CPRD OR GPRD OR Optum OR PHARMO OR HealthCore OR Danish registries

2. database:ab,ti OR (('practitioner'/exp OR practitioner) AND data AND bank) OR (('practitioner'/exp OR practitioner) AND ('database'/exp OR database)) OR ('insurance' AND ('database'/exp OR database))

3. #1 OR #2

4. 'sensitivity and specificity'/exp OR 'sensitivity and specificity'

5. specificity:ab,ti

6. 'predictive value of tests'/exp OR 'predictive value of tests'

7. (positive:ab,ti AND predictive:ab,ti AND value:ab,ti) OR (negative:ab,ti AND predictive:ab,ti AND value:ab,ti) OR (likelyhood:ab,ti AND ratio:ab,ti) OR (receiver:ab,ti AND operating:ab,ti AND characteristic:ab,ti) OR kappa:ab,ti

8. case NEAR/1 (verificat* OR valid* OR identif* OR definition* OR define* OR evaluat*)

9. 'algorithms'/exp OR algorithm

10. #4 OR #5 OR #6 OR #7 OR #8 OR #9

11. 'emphysema'/exp

12. 'chronic bronchitis'/exp

13. (obstruct* NEAR/3 (pulmonary or lung* or airway* or airflow* or bronch* or respirat*))

14. 'chronic obstructive pulmonary disease '/exp

15. 'COPD'/exp

16. 'CAPD'/exp

17. #11 OR #12 OR #13 OR #14 OR #15 OR #16

18. #3 AND #10 AND #17

Web of Science search strategy

1. (health NEAR/3 administrative) OR (administrative NEAR/3 data) OR (administrative NEAR/3 database) OR (claim NEAR/3 administrative) OR (International Classification of Diseases) OR ICD-9-CM OR ICD-10 OR (Database Management Systems) OR ("Medical Records Systems" NEAR/2 Computerized) OR "CPT" OR (Current procedural terminology) OR (computerized medical records systems) OR (electronic healthcare record) OR (computerized medical record) OR (electronic medical record) OR (automated medical record) OR (electronic patient record) OR CPRD OR GPRD OR Optum OR PHARMO OR HealthCore OR Danish registries

2. (factual databases) OR (geographic information systems) OR (national practitioner data bank) OR (insurance database)

3. #1 OR #2

4. sensitivity or "Sensitivity and Specificity"

5. specificity

6. (positive predictive value) OR (negative predictive value) OR (likelihood ratio) OR (receiver operating characteristic) OR kappa

7. ((case or cases) AND (verificat* OR valid* OR identif* OR definition* OR define* OR evaluat*))

8. algorithm

9. #4 OR #5 OR #6 OR #7 OR #8

10. emphysema

11. (chronic* NEAR/3 bronchitis*)

12. (obstruct* NEAR/3 (pulmonary or lung* or airway* or airflow* or bronch* or respirat*))

13. chronic obstructive pulmonary disease

14. (COPD)

15. (CAPD)

16. #10 OR #11 OR #12 OR #13 OR #14 OR #15

17. # 3 AND #9 AND #16

The Cochrane Library

1. (health near/3 administrative) or (administrative near/3 data) or (administrative near/3 database) or (claim near/3 administrative) or (International Classification of Diseases) or [mh "International Classification of Diseases"] or ICD-9-CM or ICD-10 or [mh "Database Management Systems"] or [mh "Medical Records Systems, Computerized"] or "CPT" or [mh "Current procedural terminology"] or (computerized medical records systems) or (electronic healthcare record) or (computerized medical record) or (electronic medical record) or (automated medical record) or (electronic patient record) or CPRD or GPRD or Optum or PHARMO or HealthCore or Danish registries

2. (factual databases) OR (geographic information systems) OR (national practitioner data bank) OR (insurance database)

3. #1 OR #2

4. sensitivity or [mh "Sensitivity and Specificity"]

5. specificity:ti,ab,kw

6. (positive predictive value) OR (negative predictive value) OR (likelihood ratio) OR (receiver operating characteristic) OR kappa

7. ((case or cases) AND (verificat* OR valid* OR identif* OR definition* OR define* OR evaluat*))

8. Algorithm OR [mh "Algorithm"]

9. #4 OR #5 OR#6 OR #7 OR #8

10. emphysema

11. (chronic* near/3 bronchitis*)

12. (obstruct* near/3 (pulmonary or lung* or airway* or airflow* or bronch* or respirat*))

13. (chronic obstructive pulmonary disease)

14. COPD

15. CAPD

16. [mh "pulmonary disease, chronic obstructive"]

17. #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16

18. # 3 AND #9 AND #17

Appendix 2

Checklist of reporting criteria for studies validating health administrative data algorithms (developed by Benchimol et al., based on the criteria published by the Standards for Reporting of Diagnostic accuracy (STARD) initiative for the accurate reporting of studies using diagnostic studies.

	YES	NO	UNCERTAIN	NOT APPLICABLE
TITLE, KEYWORDS, ABSTRACT				
Identify article as study of assessing diagnostic accuracy				
Identify article as study of administrative data				
INTRODUCTION:				
State disease identification & validation one of goals of study				
METHODS:				
Participants in validation cohort:				
Describe validation cohort (Cohort of patients to which reference standard was applied)				
• Age				
• Disease				
• Severity				
• Location/Jurisdiction				
Describe recruitment procedure of validation cohort				
• Inclusion criteria				
• Exclusion criteria				
Describe patient sampling (random, consecutive, all, etc.)				
Describe data collection				
• Who identified patients and did selection adhere to patient recruitment criteria				
• Who collected data				
• *A priori* data collection form				
• Disease classification				
• Split sample (i.e. re-validation using a separate cohort) a) Training set b) Testing set				
Test Methods:				
Describe number, training and expertise of persons reading reference standard				
If >1 person reading reference standard, quote measure of consistency (e.g. kappa)				

Blinding of interpreters of reference standard to results of classification by administrative data e.g. Chart abstractor blinded to how that chart was coded				
Statistical Methods:				
Describe methods of calculating/comparing diagnostic accuracy				
RESULTS:				
Participants:				
Report when study done, start/end dates of enrollment				
Describe number of people who satisfied inclusion/exclusion criteria				
Study flow diagram				
Test results:				
Report distribution of disease severity				
Report cross-tabulation of index tests by results of reference standard				
Estimates:				
Report at least 4 estimates of diagnostic accuracy				
Diagnostic Accuracy Measures Reported:				
• Sensitivity				
• Spec				
• PPV				
• NPV				
• Likelihood ratios				
• Kappa				
• Area under the ROC curve / c-statistic				
• Accuracy/agreement				
• Other (specify)				
Report accuracy for subgroups (e.g. age, geography, different sex, etc.)				
If PPV/NPV reported, ratio of cases/controls of validation cohort approximate prevalence of condition in the population				
Report 95% confidence intervals for each diagnostic measure				
DISCUSSION:				
Discuss the applicability of the validation findings				