HHS Public Access
Author manuscript
Pharmacoepidemiol Drug Saf. Author manuscript; available in PMC 2017 March 01.

Published in final edited form as:
Pharmacoepidemiol Drug Saf. 2016 March ; 25(3): 263–268. doi:10.1002/pds.3870.

Bayesian estimation of the accuracy of ICD-9-CM- and CPT-4-based algorithms to identify cholecystectomy procedures in administrative data without a reference standard

S. Reza Jafarzadeh[1,*], **David K. Warren**[1], **Katelin B. Nickel**[1], **Anna E. Wallace**[2], **Victoria J. Fraser**[1], and **Margaret A. Olsen**[1,3]

[1]Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA

[2]HealthCore Inc., Wilmington, DE, USA

[3]Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA

Abstract

Purpose—To estimate the accuracy of two algorithms to identify cholecystectomy procedures using International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) and Current Procedural Terminology (CPT-4) codes in administrative data.

Methods—Private insurer medical claims for 30,853 patients 18–64 years with an inpatient hospitalization between 2006 and 2010, as indicated by providers/facilities place of service in addition to room and board charges, were cross-classified according to the presence of codes for cholecystectomy. The accuracy of ICD-9-CM- and CPT-4-based algorithms was estimated using a Bayesian latent class model.

Results—The sensitivity and specificity were 0.92 [probability interval (PI): 0.92, 0.92] and 0.99 (PI: 0.97, 0.99) for ICD-9-CM-, and 0.93 (PI: 0.92, 0.93) and 0.99 (PI: 0.97, 0.99) for CPT-4-based algorithms, respectively. The parallel-joint scheme, where positivity of either algorithm was considered a positive outcome, yielded a sensitivity and specificity of 0.99 (PI: 0.99, 0.99) and 0.97 (PI: 0.95, 0.99), respectively.

Conclusions—Both ICD-9-CM- and CPT-4-based algorithms had high sensitivity to identify cholecystectomy procedures in administrative data when used individually and especially in a parallel-joint approach.

Keywords

Bayesian; cholecystectomy; latent class models; no reference standard; sensitivity; specificity

[*]Correspondence to: S.R. Jafarzadeh, Department of Medicine, Washington University School of Medicine, Campus Box 8051, 660 South Euclid Avenue, St. Louis, MO 63110, USA. rjafarza@dom.wustl.edu (S.R. Jafarzadeh)..

Conflict of interest: none declared.

This work was presented in the 2013 American College of Epidemiology Annual Meeting in Louisville, KY, USA.

INTRODUCTION

In the United States, approximately 700,000 cholecystectomy procedures are performed each year,[1] which makes it one of the most commonly performed surgical procedures. Administrative data have been used in several studies to identify cholecystectomy procedures.[2-9] These data are increasingly used for health services and outcomes research,[10-13] which is a concern due to uncertainty regarding the accuracy of diagnostic and procedural coding.[14-16] Administrative data are comprised of detailed health information including electronic insurance enrollment information, basic demographics, and claims submitted by providers and facilities for the purpose of billing and reimbursement. Given the distinct differences in data collected for billing versus research, it may be difficult to ascertain the true status of subjects with regard to an outcome or an exposure of interest when using administrative data.

To the best of our knowledge, there are no studies that assess the accuracy of different coding algorithms to identify cholecystectomy procedures in administrative data. The traditional method to validate a diagnostic algorithm is to compare it to a criterion with perfect accuracy, referred to as the reference or 'gold' standard. For administrative data, medical record (chart) review is often be used as the reference standard. However, a reference standard may not always be available for some administrative data studies due to the costs of obtaining another data source to use as a reference standard or lack of access to charts for record review. Without a reference standard, relying on a claims-based algorithm to identify subjects' status in a study population could cause misclassification of the subjects' outcome or exposure status, leading to biased estimates of prevalence or association between risk factors and an outcome.[17-19] In the absence of a reference standard, model-based methods, such as Bayesian latent class models, were developed to estimate the accuracy of diagnostic tests.[20-22] Bayesian methods allow incorporation of uncertainty as well as current knowledge regarding the parameters of interest into analysis through probability distributions, referred to as priors. Furthermore, Bayesian methods do not rely on large sample approximations for parameter estimation and provide direct probability interpretation of results from posterior distributions.[23]

The purpose of this study was to estimate the accuracy of two different algorithms to identify cholecystectomy procedures based on the two commonly used coding systems in administrative data, the International Classification of Diseases, 9th Edition, Clinical Modification (ICD-9-CM) and Current Procedural Terminology (CPT-4) codes, using Bayesian latent class models. Moreover, we provide estimates of the accuracy of the coding algorithms when used jointly in a parallel scheme.

METHODS

Data

Medical claims data from 13 Blue Cross/Blue Shield health plans contained in the HealthCore Integrated Research Database (HIRD[SM]) were used. The data included facility and professional adjudicated and paid claims for members ages 18–64 years enrolled in a non-capitated medical (hospital and physician) health plan from January 2006 through

December 2010. We restricted this study to an inpatient population because ICD-9-CM procedure codes are reported by facilities for inpatient stays, but not routinely for outpatient or ambulatory visits. Each medical claim could include unlimited CPT-4 codes and up to five ICD-9-CM procedure codes.

The inpatient stay was established based on one or more claims with an inpatient place of service assigned by a facility or a provider; consecutive dates on claims with an inpatient place of service defined the duration of the surgical episode. A room and board charge, defined as Uniform billing revenue codes in the range 0100–0249, was required during the surgical episode to meet the definition of an inpatient surgical episode. Whenever there was more than one distinct surgical episode corresponding to cholecystectomy for a patient, only the first surgical episode was included.

Algorithms

Two algorithms were considered to identify cholecystectomy procedures. The first algorithm was based on ICD-9-CM procedure codes only. In this algorithm any patient with an inpatient stay with ICD-9-CM procedure codes 51.21–51.24 was considered positive for having had a cholecystectomy procedure performed.

The second algorithm was based on CPT-4 codes only. In the CPT-4 algorithm, any inpatient stay coded by a provider with 47562, 47563, 47564, 47600, 47605, 47610, 47612, or 47620 was considered positive for having had a cholecystectomy procedure performed.

Modeling

The accuracy of a binary diagnostic test is often presented by measures such as sensitivity, the probability of a positive result in the population with the condition of interest, and specificity, the probability of a negative result in the population without the condition of interest. Without a ‘perfect’ method of ascertainment or a reference standard, the accuracy of coding algorithms is ‘latent’,(i.e., the accuracy cannot be directly observed, but can be estimated from data).[24,25] We used a modified version of the Bayesian latent class model of ‘two tests, one population’, as discussed in Joseph et al 1995,[20] which does not rely on information from a reference standard to estimate the accuracy measures of interest, i.e. sensitivity and specificity. Hui and Walter 1980 provided methods to estimate the accuracy of two tests without a reference standard.[26] The problem in estimating the five parameter of interest, two sensitivities, two specificities and prevalence, is that there are only three independent cells (degrees of freedom) available in a cross-classified table of test results, which leads to statistical non-identifiability.[27] Therefore, the problem is unsolvable using frequentist methodology without additional assumptions. Hui and Walter required a second population with distinct prevalence in order to increase the degrees of freedom required to estimate the five parameters of interest in addition to the prevalence in the second population. Bayesian approach is able to tackle this problem and lack of identifiability (or weak identifiability) in the ‘two tests, one-population’ scenario by specifying probability distributions on the parameters of interest.[27]

In a Bayesian approach, uncertainties regarding the accuracy of coding algorithms are represented by probability distributions. These probability distributions, commonly referred to priors, are combined with a likelihood function of data to obtain updated probability distributions, referred to as posterior distributions, for all parameters of interest. A detailed description of the model is in the Appendix. Briefly, the results for the two algorithms were cross-classified, where it was assumed that the observed cross-classified results have a multinomial sampling distribution. The multinomial probabilities associated with the cross-classified results can be described by the true-positive (sensitivity), false-positive (‘1 – sensitivity’), true-negative (specificity) and false-negative (‘1 – specificity’) probabilities, and the probability of the condition in the population.[20,22] It was assumed that the results of the two algorithms are independent of each other conditional on the true cholecystectomy status. The conditional independence assumption is reasonable since CPT-4 and ICD-9-CM procedure codes are coded independently by separate entities (i.e., facility medical coders for ICD-9-CM procedure codes and clinicians or their staff for CPT-4 codes). For situations where algorithms may be dependent, a model that incorporates conditional dependence could be used.[22,28]

We then estimated the sensitivity and specificity of the two algorithms. A parallel-testing scheme is often implemented for screening or diagnostic algorithms in order to improve or gain increased sensitivity. In the parallel-testing scheme, having a positive result from either of the two algorithms would be considered as a positive outcome. Therefore, we also estimated the sensitivity and specificity of the two algorithms when they were used in parallel combination.

Priors

We used BetaBuster software (http://www.epi.ucdavis.edu/diagnostictests/betabuster.html) to obtain a unique beta prior corresponding to our prior belief. The BetaBuster requires two quantities to obtain a unique beta distribution. We first determined our best guess for the most likely value for a parameter, for example, sensitivity of the ICD-9-CM-based algorithm, and then our best guess that the most likely value for the parameter is more (or less) than a specific value. BetaBuster uses the most likely value as the mode and the other quantity as the 5th (or 95th) percentile of the corresponding beta distribution. For example, we assumed that the most likely value (mode) for the sensitivity of the ICD-9-CM-based algorithm is 0.95, and we are 95% sure that the mode is more than 0.80, which corresponds to the Beta(21.20, 2.06) distribution. The elicited priors used for the Bayesian model parameters are summarized in Table 1.

The model described in the previous section is non-identifiable; however, achieving identifiability is not mandatory in Bayesian analysis with informative priors.[27] Therefore, priors were specified to be consistent with the data, where implausible (or extremely unlikely) and non-informative priors were not considered. Sensitivity analyses for our proposed model were performed using two sets of less informative priors (Table 1) to assess how posterior estimates would be affected by our choice of priors.

Computation

The Bayesian modeling was performed in JAGS version 3.4.0[29] through ‘rjags’[30] library in R[31] software.[31] All inferences were based on 100,000 iterations thinned from 100,000,000 after a burn-in of 10,000,000 iterations. Geweke's statistic was used to assess lack of convergence for the Markov chains through 'coda'[32] library in R.

RESULTS

The study population included 30,853 patients with a cholecystectomy coded during an inpatient hospital stay. Table 2 presents the cross-classified results for the ICD-9-CM- and CPT-4-based algorithms for the study population.

The posterior median and 95% probability interval for the sensitivities and specificities of the two algorithms when used separately and also when used in parallel are summarized in Table 3. Both algorithms had high accuracy to identify cholecystectomy procedures in this data. When used individually, it was estimated that the false negative proportions were about 8% and 7% for the ICD-9-CM- and CPT-4-based algorithms, respectively. The highest sensitivity of detection was obtained when the algorithms were used jointly in the parallel scheme (0.99, probability interval [PI]: 0.99, 0.99; Table 3). Sensitivity analysis with separate sets of less informative priors produced similar results (Table 3).

DISCUSSION

Our study suggests that both the ICD-9-CM- and CPT-4-based algorithms have high accuracy, especially when used in parallel combination, to identify cholecystectomy procedures in administrative data. Several studies using administrative data have relied on ICD-9-CM procedure codes alone to identify cholecystectomy procedures.[4,5,8] Some investigators have required additional ICD-9-CM diagnosis codes for conditions such as gallbladder disease along with ICD-9-CM procedure codes to identify cholecystectomy.[2,3,6,7,9] This approach, commonly referred to as serial-testing, could potentially lead to a higher specificity to detect cholecystectomy at the expense of a loss in sensitivity. In some studies, a reference standard was used to validate the detection algorithm only for subjects who had ICD-9-CM procedure codes for cholecystectomy;[4,8] consequently, unless adjusted for in the analysis or the design of the study, this could lead to partial verification bias, also referred to as workup bias.[33] Weinhandl and Gilbertson showed that the commonly used case-control design, where the validation study population includes a random sample of the subjects without any code in addition to all the subjects that are positive due to an algorithm, could potentially result in biased estimates of the accuracy and invalid findings because estimates for specificity and (consequently) predictive values vary considerably with the sampling rate.[34]

In our study, we did not rely on a reference standard to verify the true classification status of subjects for each algorithm. Instead, we relied on a model-based approach to estimate the accuracy of the algorithms through Bayesian modeling, where current knowledge and uncertainty about the accuracy of the algorithms were represented by probability distributions. Model-based estimation of the accuracy of algorithms in administrative data has previously been used to identify conditions or procedures such as osteoarthritis,[19] systemic lupus erythematosus,[35] autoimmune rheumatic disease[36] and screening colonoscopy[37] among others. An important feature of our study is that we used a national insurer database that represents multiple hospitals and physician groups of various sizes in both academic and private settings, which strengthens its generalizability.

In general, inaccuracy in administrative data can occur in several ways. For example, coding for some procedures may be complicated and subject to interpretation. Untimely documentation by clinicians could lead to inaccuracy in medical records if coders need to complete their coding documentation before operating notes are complete.[38] Some hospitals may use commercially available coding software rather than medical coders, likely resulting in some degree of coding error.

It is possible that we had missing records on CPT-4 codes from providers in our data potentially due to denial of provider claims by private insurers or incomplete enrollment information for patients with dual insurance; consequently, we may not have had complete information on CPT-4 codes from providers, which would have resulted in a lower sensitivity for the CPT-4-based algorithm. The number of ICD-9-CM procedure codes on each hospital claim in our data was limited to five. The primary purpose of administrative data is reimbursement of healthcare costs; thus, multiple diagnosis or procedure codes are likely to be listed in order to maximize remuneration. It is possible that ICD-9-CM procedure codes could have been dropped if there were other conditions associated with higher reimbursement occurring during the same inpatient admission.[12] Another option is that ICD-9-CM codes may be listed temporally, so some codes may be missed if a procedure was performed relatively late into the admission, especially for long hospitalizations. However, possible dropping of ICD-9-CM procedure codes did not appear to be a problem in our data. Only about 0.4% of patients, who were assigned a CPT-4 code for cholecystectomy by providers, had a facility claim that used all five fields allocated for ICD-9-CM procedure codes during their surgical admission without having an ICD-9-CM procedure code for cholecystectomy on their claims.

Our study population was based on inpatient hospitalizations; therefore, the parallel combination would not be applicable for ambulatory procedures since CPT-4 codes are assigned by both facilities and providers in ambulatory settings. The parallel combination would also not be applicable for databases containing only hospital billing data, as those data would be restricted to only ICD-9-CM procedure codes. Moreover, the parallel approach, for the purpose of maximizing sensitivity, may not be the most reasonable choice in all settings or for other procedures, for example when identifying true-positives does not outweigh consequences associated with including more false-positives in the study population.

Even though we have used administrative data from a large private insurer that cover regions in the Southeast, Mid-Atlantic, Eastern, Central, and Western regions of the United States, it should be noted that the assumption of constant sensitivity and specificity across populations may not hold unless the factors contributing to heterogeneity of populations could be measured and controlled. Therefore, sensitivity and specificity of our algorithms may not be the same when they are applied to populations and administrative datasets with distinctly different characteristics.[39] The two algorithms that we evaluated are coded independently by distinct entities (ICD-9-CM procedure codes by facilities and CPT-4 codes by providers) and are considered being inherently independent. There are uncommon instances where these could be coded by the same group in a hospital, for example by a hospital-employed surgeon. However, we avoided estimating two additional parameters for conditional covariances between the sensitivities and specificities of the algorithms because there is no extra information in data and the posterior estimates for conditional covariances would be the same as priors on these covariances.

The findings of this study suggest that both the ICD-9-CM- and CPT-4-based algorithms, especially in parallel combination, have high accuracy to identify cholecystectomy procedures in inpatient hospital populations using administrative data. Using multiple algorithms to increase the accuracy of detection is a common practice; however, the gain by parallel interpretation of these two algorithms is marginal since both algorithms are estimated to be highly accurate. In practice, whether an investigator uses multiple algorithms in parallel-, serial- or sequential-schemes or even individually depends on a specific setting, the accuracy of each individual algorithm, and the trade-off between sensitivity and specificity of the chosen scheme. Further, the study illustrates use of Bayesian latent class model to estimate accuracy of coding algorithms in administrative data.

ACKNOWLEDGMENTS

Research reported in this study was supported by the Prevention Epicenters Program for the Prevention of Healthcare Associated Infections from the Center for Disease Control and Prevention under award number U54 CK000162 and cooperative agreement FOA# CK11-001.

Sponsors: This study was supported by the Agency for Healthcare Research and Quality (grant 5R01HS019713) and the Prevention Epicenters Program from the Centers for Disease Control and Prevention (grant U54 CK000162).

APPENDIX

The resulting data after applying the ICD-9-CM- and CPT-4-based algorithms, denoted as algorithms 1 and 2, respectively, are cross-classified. It is assumed that the results of the two algorithms for a given patient are independent, conditional on the true status of each patient with regard to whether a cholecystectomy procedure was performed or not.

Let $X = (X_{11}, X_{10}, X_{01}, X_{00})$ be the vector of results for the two algorithm, where X_{11} is the number of patients that are positive for both algorithm 1 and 2, X_{10} (X_{01}) is the number of patients that are positive (negative) for algorithm 1 (algorithm 2), and X_{00} is the number of patients that are negative for both algorithms 1 and 2. The observed cross-classified results are modeled as multinomial:

$$X\tilde{}\text{multinomial}\left(N, \left(p_{11}, p_{10}, p_{01}, p_{00}\right)\right),$$

where N is the sampled population and p's are the multinomial probabilities corresponding to observed data, and are given by

$$
\begin{aligned}
p_{11}= & \pi Se_1 Se_2 + (1-\pi)(1-Sp_1)(1-Sp_2), \\
p_{10}= & \pi Se_1 (1-Se_2) + (1-\pi)(1-Sp_1) Sp_2, \\
p_{01}= & \pi (1-Se_1) Se_2 + (1-\pi) Sp_1 (1-Sp_2), \\
p_{00}= & \pi (1-Se_1)(1-Se_2) + (1-\pi) Sp_1 Sp_2,
\end{aligned}
$$

where π is the probability of the condition of interest in the population, and Se and Sp are the corresponding sensitivities and specificities of algorithms 1 and 2. For example, p_{11} is calculated as p_{11} = Pr(T1+,T2+) = Pr(T1+,T2+|D+)Pr(D+) + Pr(T1+,T2+|D−)Pr(D−) = $\pi Se_1 Se_2 + (1 - \pi)(1 - Sp_1)(1 - Sp_2)$, where T refers to an algorithm and D refers to true cholecystectomy status.

Furthermore, the sensitivity and specificity of the parallel-joint scheme are $Se_{Parallel} = Se_1 + Se_2 - Se_1 Se_2$ and $Sp_{Parallel} = Sp_1 Sp_2$, respectively. JAGS version 3.4.0 code to run the model:

```
model{
# likelihood
X[1:4] ~ dmulti(p[1:4], N)
p[1] <- pi * Se1 * Se2 + (1 - pi) * (1 - Sp1) * (1 - Sp2)
p[2] <- pi * Se1 * (1 - Se2) + (1 - pi) * (1 - Sp1) * Sp2
p[3] <- pi * (1 - Se1) * Se2 + (1 - pi) * Sp1 * (1 - Sp2)
p[4] <- pi * (1 - Se1) * (1 - Se2) + (1 - pi) * Sp1 * Sp2
Se.Parallel <- Se1 + Se2 - Se1 * Se2
Sp.Parallel <- Sp1 * Sp2
# priors
pi ~ dbeta(, )
Se1 ~ dbeta(, )
Sp1 ~ dbeta(, )
Se2 ~ dbeta(, )
Sp2 ~ dbeta(, )
}
```

If X_{00} is not available, data $X = (X_{11}, X_{10}, X_{01})$ can be modeled as

$$X\tilde{}\mathrm{multinomial}\left(N, \left(p_{11}, p_{10}, p_{01}\right)\right)$$

where

$$\begin{aligned} p_{11}= & \pi Se_1 Se_2 + (1-\pi)(1-Sp_1)(1-Sp_2), \\ p_{10}= & \pi Se_1 (1-Se_2) + (1-\pi)(1-Sp_1) Sp_2, \\ p_{01}= & 1-(p_{11}+p_{10}) \end{aligned}$$

The second scenario is more common in administrative data because often study data are extracted from complex and multi-layered administrative databases by querying for specific diagnosis or procedure codes; however, such data may not be accessible on all patients, especially patients who don't have these specific codes.

REFERENCES

1. Shaffer EA. Epidemiology and risk factors for gallstone disease: has the paradigm changed in the 21st century? Curr Gastroenterol Rep. 2005; 7:132–140. [PubMed: 15802102]
2. Steiner CA, Bass EB, Talamini MA, Pitt HA, Steinberg EP. Surgical rates and operative mortality for open and laparoscopic cholecystectomy in Maryland. N Engl J Med. 1994; 330:403–408. doi: 10.1056/NEJM199402103300607. [PubMed: 8284007]
3. Laycock WS, Siewers AE, Birkmeyer CM, Wennberg DE, Birkmeyer JD. Variation in the use of laparoscopic cholecystectomy for elderly patients with acute cholecystitis. Arch Surg Chic Ill 1960. 2000; 135:457–462.
4. Ahmed HU, Smith JB, Rudderow DJ, Longo WE, Virgo KS, Johnson FE. Cholecystectomy in patients with previous spinal cord injury. Am J Surg. 2002; 184:452–459. [PubMed: 12433613]
5. Livingston EH, Rege RV. A nationwide study of conversion from laparoscopic to open cholecystectomy. Am J Surg. 2004; 188:205–211. doi:10.1016/j.amjsurg.2004.06.013. [PubMed: 15450821]
6. Varela JE, Nguyen NT. Disparities in access to basic laparoscopic surgery at U.S. academic medical centers. Surg Endosc. 2011; 25:1209–1214. doi:10.1007/s00464-010-1345-y. [PubMed: 20848139]
7. Tiwari MM, Reynoso JF, High R, Tsang AW, Oleynikov D. Safety, efficacy, and cost-effectiveness of common laparoscopic procedures. Surg Endosc. 2011; 25:1127–1135. doi:10.1007/s00464-010-1328-z. [PubMed: 20927546]
8. Allam E, Patel A, Lewis G, et al. Cholecystectomy in patients with prior ventriculoperitoneal shunts. Am J Surg. 2011; 201:503–507. doi:10.1016/j.amjsurg.2010.05.006. [PubMed: 21421103]
9. Agabiti N, Stafoggia M, Davoli M, Fusco D, Barone AP, Perucci CA. Thirty-day complications after laparoscopic or open cholecystectomy: a population-based cohort study in Italy. BMJ Open. 2013; 3 doi:10.1136/bmjopen-2012-001943.
10. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005; 58:323–337. doi:10.1016/j.jclinepi.2004.10.012. [PubMed: 15862718]
11. Wunsch H, Harrison DA, Rowan K. Health services research in critical care using administrative data. J Crit Care. 2005; 20:264–269. doi:10.1016/j.jcrc.2005.08.002. [PubMed: 16253796]
12. Jhung MA, Banerjee SN. Administrative coding data and health care-associated infections. Clin Infect Dis Off Publ Infect Dis Soc Am. 2009; 49:949–955. doi:10.1086/605086.
13. Miriovsky BJ, Shulman LN, Abernethy AP. Importance of health information technology, electronic health records, and continuously aggregating data to comparative effectiveness research and learning health care. J Clin Oncol Off J Am Soc Clin Oncol. 2012; 30:4243–4248. doi: 10.1200/JCO.2012.42.8011.

14. Tamblyn R, Lavoie G, Petrella L, Monette J. The use of prescription claims databases in pharmacoepidemiological research: the accuracy and comprehensiveness of the prescription claims database in Québec. J Clin Epidemiol. 1995; 48:999–1009. [PubMed: 7775999]
15. De Coster C, Quan H, Finlayson A, et al. Identifying priorities in methodological research using ICD-9-CM and ICD-10 administrative data: report from an international consortium. BMC Health Serv Res. 2006; 6:77. doi:10.1186/1472-6963-6-77. [PubMed: 16776836]
16. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. J Clin Epidemiol. 2011; 64:1054–1059. doi:10.1016/j.jclinepi.2011.01.001. [PubMed: 21474278]
17. Copeland KT, Checkoway H, McMichael AJ, Holbrook RH. Bias due to misclassification in the estimation of relative risk. Am J Epidemiol. 1977; 105:488–495. [PubMed: 871121]
18. Walter SD, Irwig LM. Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. J Clin Epidemiol. 1988; 41:923–937. doi:10.1016/0895-4356(88)90110-2. [PubMed: 3054000]
19. Ladouceur M, Rahme E, Pineau CA, Joseph L. Robustness of prevalence estimates derived from misclassified data from administrative databases. Biometrics. 2007; 63:272–279. doi:10.1111/j.1541-0420.2006.00665.x. [PubMed: 17447953]
20. Joseph L, Gyorkos TW, Coupal L. Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. Am J Epidemiol. 1995; 141:263–272. [PubMed: 7840100]
21. Enøe C, Georgiadis MP, Johnson WO. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. Prev Vet Med. 2000; 45:61–81. [PubMed: 10802334]
22. Branscum AJ, Gardner IA, Johnson WO. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev Vet Med. 2005; 68:145–163. doi:10.1016/j.prevetmed.2004.12.005. [PubMed: 15820113]
23. Christensen, R.; Johnson, WO.; Branscum, AJ.; Hanson, TE. Bayesian Ideas and Data Analysis: an Introduction for Scientists and Statisticians. 1st ed.. CRC Press; 2010.
24. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM. A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. J Clin Epidemiol. 2009; 62:797–806. doi:10.1016/j.jclinepi.2009.02.005. [PubMed: 19447581]
25. van Smeden M, Naaktgeboren CA, Reitsma JB, Moons KGM, de Groot JAH. Latent class models in diagnostic studies when there is no reference standard--a systematic review. Am J Epidemiol. 2014; 179:423–431. doi:10.1093/aje/kwt286. [PubMed: 24272278]
26. Hui SL, Walter SD. Estimating the error rates of diagnostic tests. Biometrics. 1980; 36:167–171. [PubMed: 7370371]
27. Jones G, Johnson WO, Hanson TE, Christensen R. Identifiability of models for multiple diagnostic testing in the absence of a gold standard. Biometrics. 2010; 66:855–863. doi:10.1111/j.1541-0420.2009.01330.x. [PubMed: 19764953]
28. Johnson WO, Gastwirth JL, Pearson LM. Screening without a “gold standard”: the Hui-Walter paradigm revisited. Am J Epidemiol. 2001; 153:921–924. [PubMed: 11323324]
29. Plummer, M. JAGS version 3.4.0 user manual. International Agency for Research on Cancer; Lyon, France: 2013.
30. Plummer, M. rjags: Bayesian graphical models using MCMC. R package version 3-12. 2014. http://cran.r-project.org/package=rjags
31. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2014. Available at: http://www.r-project.org
32. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. R News. 2006; 6:7–11.
33. Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics. 1983; 39:207–215. [PubMed: 6871349]
34. Weinhandl ED, Gilbertson DT. Invalid sampling in a validation study of diagnosis codes for infection. J Clin Epidemiol. 2011; 64:574–575. doi:10.1016/j.jclinepi.2010.11.006. [PubMed: 21324646]

35. Bernatsky S, Joseph L, Pineau CA, Tamblyn R, Feldman DE, Clarke AE. A population-based assessment of systemic lupus erythematosus incidence and prevalence--results and implications of using administrative data for epidemiological studies. Rheumatol Oxf Engl. 2007; 46:1814–1818. doi:10.1093/rheumatology/kem233.
36. Bernatsky S, Lix L, Hanly JG, et al. Surveillance of systemic autoimmune rheumatic diseases using administrative data. Rheumatol Int. 2011; 31:549–554. doi:10.1007/s00296-010-1591-2. [PubMed: 20665025]
37. Sewitch MJ, Jiang M, Joseph L, Hilsden RJ, Bitton A. Developing model-based algorithms to identify screening colonoscopies using administrative health databases. BMC Med Inform Decis Mak. 2013; 13:45. doi:10.1186/1472-6947-13-45. [PubMed: 23574795]
38. Wang MC, Laud PW, Macias M, Nattinger AB. Strengths and limitations of International Classification of Disease Ninth Revision Clinical Modification codes in defining cervical spine surgery. Spine. 2011; 36:E38–44. doi:10.1097/BRS.0b013e3181d273f6. [PubMed: 20975624]
39. Johnson WO, Gardner IA, Metoyer CN, Branscum AJ. On the interpretation of test sensitivity in the two-test two-population problem: assumptions matter. Prev Vet Med. 2009; 91:116–121. doi:10.1016/j.prevetmed.2009.06.006. [PubMed: 19651450]

Author Manuscript

Author Manuscript

Author Manuscript

KEY POINTS

- The ICD-9-CM- and CPT-4 based algorithms have high accuracy to identify cholecystectomy procedures in administrative data. This has not been demonstrated before despite the use of administrative data to identify cholecystectomy in several past studies.
- Whenever possible, the use of ICD-9-CM- and CPT-4-based algorithms used in parallel combination is recommended to achieve the highest sensitivity to identify cholecystectomy procedures in administrative data.

Table 1

Elicited priors for the sensitivities and specificities of the ICD-9-CM- and CPT-4-based algorithms.

Algorithm	Parameter	Parameter description and corresponding prior	Alternative prior for sensitivity analysis 1	Alternative prior for sensitivity analysis 2
	Pr	Mode = 0.97, 95% sure that mode > 0.90; Beta(53.58, 2.63)	Mode = 0.97, 95% sure that mode > 0.90; Beta(53.58, 2.63)	Mode = 0.97, 95% sure that mode > 0.90; Beta(53.58, 2.63)
ICD-9-CM-based	Se	Mode = 0.95, 95% sure that mode > 0.80; Beta(21.20, 2.06)	Mode = 0.95, 95% sure that mode > 0.65; Beta(8.45, 1.39)	Mode = 0.80, 95% sure that mode > 0.60; Beta(14.8442, 4.4611)
	Sp	Mode = 0.99, 95% sure that mode > 0.97; Beta(212.12, 3.13)	Mode = 0.99, 95% sure that mode > 0.95; Beta(88.28, 1.88)	Mode = 0.99, 95% sure that mode > 0.95; Beta(88.28, 1.88)
CPT-4-based	Se	Mode = 0.95, 95% sure that mode > 0.80; Beta(21.20, 2.06)	Mode = 0.95, 95% sure that mode > 0.65; Beta(8.45, 1.39)	Mode = 0.80, 95% sure that mode > 0.60; Beta(14.8442, 4.4611)
	Sp	Mode = 0.99, 95% sure that mode > 0.97; Beta(212.21, 3.13)	Mode = 0.99, 95% sure that mode > 0.95; Beta(88.28, 1.88)	Mode = 0.99, 95% sure that mode > 0.95; Beta(88.28, 1.88)

Pr = probability of the condition in the population; Se = sensitivity; Sp = specificity.

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Cross-classified results of the ICD-9-CM- and CPT-4-based algorithms for cholecystectomy procedures.

	ICD-9-CM-based	CPT-4-based	Frequency
	+	+	26 446 (85.7%)
	+	−	2 091 (6.8%)
	−	+	2 316 (7.5%)
Total			30 853

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Posterior medians and 95% probability intervals for the accuracy of the ICD-9-CM- and CPT-4-based algorithms when used separately and jointly in parallel.

Algorithm	Parameter	Posterior median (95% PI)	Posterior median (95% PI) using priors in sensitivity analysis 1	Posterior median (95% PI) using priors in sensitivity analysis 2
ICD-9-CM-based	Se	0.92 (0.92, 0.92)	0.92 (0.92, 0.92)	0.92 (0.92, 0.92)
	Sp	0.99 (0.97, 0.99)	0.98 (0.94, 0.99)	0.98 (0.94, 0.99)
CPT-4-based	Se	0.93 (0.92, 0.93)	0.93 (0.92, 0.93)	0.93 (0.92, 0.93)
	Sp	0.99 (0.97, 0.99)	0.98 (0.94, 0.99)	0.98 (0.94, 0.99)
Parallel-joint	Se	0.99 (0.99, 0.99)	0.99 (0.99, 0.99)	0.99 (0.99, 0.99)
	Sp	0.97 (0.95, 0.99)	0.96 (0.91, 0.99)	0.96 (0.91, 0.99)

PI = probability interval; Se = sensitivity; Sp = specificity.