Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2016 December ; 25(12): 1368–1374. doi:10.1002/pds.4116.

Development and Validation of Case-Finding Algorithms for the Identification of Patients with ANCA-Associated Vasculitis in Large Healthcare Administrative Databases

Antoine G. Sreih¹, Narender Annapureddy², Jason Springer³, George Casey⁴, Kevin Byram², Andy Cruz⁵, Maya Estephan³, Vince Frangiosa⁵, Michael D. George¹, Mei Liu⁶, Adam Parker³, Sapna Sangani¹, Rebecca Sharim¹, and Peter A. Merkel^{1,7} for the Vasculitis Patient-Powered Research Network

¹Vasculitis Center, Division of Rheumatology, University of Pennsylvania, Philadelphia, PA

²Division of Rheumatology, Vanderbilt University Medical Center, Nashville, TN

³Division of Rheumatology, University of Kansas Medical Center, Kansas City, KS

⁴The Vasculitis Foundation, Kansas City, MO

⁵Penn Medicine Academic Computing Services, The University of Pennsylvania, Philadelphia, PA

⁶Division of Medical Informatics, Department of Internal Medicine, University of Kansas Medical Center, Kansas City, KS

⁷Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA

Abstract

Purpose—To develop and validate case-finding algorithms for granulomatosis with polyangiitis (Wegener's, GPA), microscopic polyangiitis (MPA), and eosinophilic granulomatosis with polyangiitis (Churg-Strauss, EGPA).

Methods—250 patients per disease were randomly selected from 2 large healthcare systems using the International Classification of Diseases version 9 (ICD9) codes for GPA/EGPA (446.4) and MPA (446.0). 16 case-finding algorithms were constructed using a combination of ICD9 code, encounter type (inpatient or outpatient), physician specialty, use of immunosuppressive medications, and the anti-neutrophil cytoplasmic antibody (ANCA) type. Algorithms with the highest average positive predictive value (PPV) were validated in a third healthcare system.

Results—An algorithm excluding patients with eosinophilia *or* asthma and including the encounter type and physician specialty had the highest PPV for GPA (92.4%). An algorithm including patients with eosinophilia *and* asthma and the physician specialty had the highest PPV for EGPA (100%). An algorithm including patients with one of the following diagnoses: alveolar hemorrhage, interstitial lung disease, glomerulonephritis, acute or chronic kidney disease, the encounter type, physician specialty, and immunosuppressive medications had the highest PPV for MPA (76.2%). When validated in a third healthcare system, these algorithms had high PPV

Conflict of interest: The authors have declared that no conflict of interest exists.

(85.9% for GPA, 85.7% for EGPA, and 61.5% for MPA). Adding the ANCA type increased the PPV to 94.4%, 100%, and 81.2% for GPA, EGPA, and MPA respectively.

Conclusion—Case-finding algorithms accurately identify patients with GPA, EGPA, and MPA in administrative databases. These algorithms can be used to assemble population-based cohorts and facilitate future research in epidemiology, drug safety, and comparative effectiveness.

Keywords

ANCA; granulomatosis with polyangiitis; eosinophilic granulomatosis with polyangiitis; microscopic polyangiitis; computable phenotypes; vasculitis

INTRODUCTION

Anti-neutrophil cytoplasmic antibodies (ANCA)-associated vasculitis (AAV) is a group of autoimmune inflammatory diseases characterized by vascular inflammation and multisystem organ damage (1). AAV is comprised of three diagnoses: granulomatosis with polyangiitis (Wegener's, GPA), eosinophilic granulomatosis with polyangiitis (EGPA, Churg-Strauss), and microscopic polyangiitis (MPA). These conditions are variably associated with the presence of ANCA. AAV is rare with a combined prevalence of 46–184/million and are associated with high morbidity and mortality (1, 2). Comparative effectiveness and drug safety studies in AAV have been hindered by the rarity of the disease and the inability to establish a homogenous cohort of patients with these rare conditions.

The International Classification of Diseases versions 9 and 10 (ICD9 and ICD10) codes have been used to identify patients with AAV in large administrative databases (3, 4). However, ICD codes lack accuracy for the identification of patients with AAV and other clinical conditions (5–7). For example, both GPA and EGPA, two pathologically and phenotypically distinct diseases, share the same ICD9 code (446.4). An identified cohort of patients using ICD9 code 446.4 may result in a heterogeneous group of patients with both GPA and EGPA. Similarly, the ICD9 code 446.0 will identify patients with MPA and polyarteritis nodosa (PAN), two pathologically and clinically different diseases (8). In addition, healthcare providers regularly use ICD codes in the course of clinical evaluations when the condition has not yet been confirmed. This potential for mislabeling, early in the process of the disease evaluation, frequently results in inaccuracy of ICD codes.

Validated case-finding algorithms beyond exclusively using ICD codes are needed to accurately identify patients with AAV. These algorithms can be used to assemble a population-based cohort and facilitate future research in pharmacoepidemiology, healthcare economics, comparative effectiveness, drug safety, and disease outcomes in patients with AAV. There have been no case-finding algorithms developed or validated in AAV except for EGPA (9). The objective of this study was to develop and validate case-finding algorithms for AAV through data-driven strategies that utilized large healthcare administrative databases.

METHODS

Population and setting

The study was performed in 3 large healthcare systems: the University of Pennsylvania Health System (System 1), the Vanderbilt University Health System (System 2), and the University of Kansas Medical Center (System 3). The first 2 healthcare systems (Systems 1 and 2) were the “algorithm-deriving” systems used to develop and identify case-finding algorithms with the highest average positive predictive value (PPV). The third healthcare system (System 3) was the “algorithm-validating” system used to validate the derived algorithms with the highest average PPV. 125 randomly-selected patients seen between January 2000 and January 2015 with at least one ICD9 code 446.4 or 446.0 were identified in each algorithm-deriving healthcare system (Table 1). Given that GPA and EGPA share the same ICD9 code (446.4), there was an overlap of patients between those two samples. The Institutional Review Board of each healthcare system approved the study.

Diagnosis confirmation

The diagnoses of GPA, EGPA, and MPA were confirmed by chart review by applying the American College of Rheumatology (ACR) classification criteria or the revised International Chapel Hill Consensus Conference (ICHCC) definitions for GPA, EGPA, and MPA (8, 10). Of the 125 charts reviewed, patients who fulfilled the ACR classification criteria or the revised ICHCC definitions for GPA, EGPA, and MPA were considered to have the disease and those who did not were considered not to have the disease. Patients who had insufficient data in their medical records to confirm or refute the diagnosis were excluded from the analysis. Three different reviewers (MG, RS, and SS) performed chart review at System 1, 1 reviewer (KB) at System 2, and 2 reviewers (ME and AP) at System 3. Reviewers were instructed about and trained in the classification criteria of each type of AAV at the beginning of the study. 30 randomly-selected charts were examined by a vasculitis specialist at each of the healthcare systems (AS in System 1, NA in System 2, and JS in System 3) to calculate the interrater reliability at each site. Patients who had insufficient data to confirm or refute the diagnosis were excluded from the final analysis.

Algorithms development

Case-finding algorithms were constructed using a combination of variables including ICD9 codes (446.4 for GPA/EGPA and 446.0 for MPA), the encounter type (either “inpatient” defined as an ICD9 code submitted on 3 consecutive days or “outpatient” defined as two ICD9 codes submitted 3 months apart), the specialty of the physician involved in the care (a Rheumatologist, Allergist/Immunologist, Nephrologist, Pulmonologist, or Otorhinolaryngologist), the use of immunosuppressive medications (see list in Supplementary Table 6), and the pattern of ANCA [cytoplasmic (cANCA) or perinuclear (pANCA)]. 16 case-finding algorithms were created for GPA, EGPA, and MPA based on clinical experience and expert opinion of the investigators and the PPV and sensitivity for the constructed algorithms in each algorithm-deriving healthcare system were calculated (Figures 1, 2, and 3). Given that healthcare systems occasionally record test results in an unsearchable format such as electronically scanned-in files or note-documented laboratory results, algorithms were developed with and without the ANCA pattern. The ANCA pattern

was chosen in lieu of tests for the specific ANCA antigens proteinase 3 (PR3) or myeloperoxidase (MPO) because some healthcare organizations continue to document only the ANCA pattern without the specific ANCA serotypes. Algorithms with the highest average PPV for GPA, EGPA, and MPA from both algorithm-deriving healthcare systems were identified. In cases in which more than one algorithm were found to have a high average PPV, the algorithm with the higher sensitivity or the least number of variables was chosen to be validated. Given that GPA and EGPA share the same ICD9 code (446.4) patients who ever had ICD9 codes 283.3 (eosinophilia) or 493.X (asthma) were excluded when testing algorithms for GPA. Alternatively, patients who ever had ICD9 codes 283.3 (eosinophilia) or 493.X (asthma) were included when testing algorithms for EGPA. Similarly, because MPA and polyarteritis nodosa (PAN) share the same ICD9 code (446.0), patients who ever had ICD9 codes 770.3 (alveolar hemorrhage) or 516.34 (interstitial lung disease) or 583.X (glomerulonephritis) or 584.X (acute kidney failure) or 585.X (chronic kidney disease) were included when testing for algorithms for MPA to limit the number of patients with PAN in the sample.

Algorithm validation

Algorithms identified as having the highest average PPV for GPA, EGPA, and MPA from both algorithm-deriving healthcare systems were selected for validation in System 3. The medical records of patients identified by these algorithms were reviewed and the diagnosis confirmed using the methods detailed above. The percentage of patients with a confirmed diagnosis of GPA, EGPA, and MPA (true PPV) was calculated.

Statistical analysis

The PPV and sensitivity for each constructed algorithm were calculated using a 2×2 contingency table. The average PPV for each constructed algorithm from both algorithm-deriving healthcare systems was determined by dividing the sum of both PPVs by 2. Scatter plots were used to illustrate the relationship between PPV and sensitivity. The inter-operator reliability was measured using the Kappa statistic. STATA version 14.0 was used to perform statistical analyses.

RESULTS

Table 1 shows the total number of patients queried in each algorithm-deriving healthcare system and the total number of patients with at least one of the ICD9 codes: 446.4 or 446.0. 486 of 3,161,862 patients had at least one ICD9 code 446.4 at System 1 and 559 out of 2,443,098 patients at System 2. Fewer patients had at least one ICD9 code 446.0 (221 and 385 patients at Systems 1 and 2, respectively). The medical records of 125 randomly-selected patients from the number of patients with at least one ICD9 code per type of AAV per algorithm-deriving healthcare system were reviewed. The interrater reliability (Kappa $\pm$ standard error) between reviewers at System 1 was 0.824 ± 0.075 , at System 2 was 0.930 ± 0.075 , and at System 3 was 0.839 ± 0.089 .

Algorithms for identifying granulomatosis with polyangiitis

Of the 125 charts reviewed per algorithm-deriving healthcare system, 97 patients from System 1 and 98 patients from System 2 had sufficient data in their medical records to confirm or refute the diagnosis of GPA and were included in the analysis. 45/97 (46.4%) patients had a confirmed diagnosis of GPA at System 1 and 66/98 (67.3%) patients at System 2 based on the ACR criteria or the revised ICHCC definitions. Figure 1 shows the PPV and sensitivity of the 16 constructed case-finding algorithms for GPA. The PPV and sensitivity of the algorithms in each healthcare system are detailed in Supplementary Table 2. An algorithm excluding patients with eosinophilia *or* asthma and including the encounter type and physician specialty had the highest average positive predictive value (PPV: 92.9% at System 1 and 92.1 % at System 2). Adding the cANCA pattern significantly increased the PPV across all algorithms (up to 100%) but decreased the sensitivity.

94 patients out of 2,151,960 were identified at the algorithm-verifying healthcare system (System 3) using the algorithm with the highest average PPV listed above. 23 patients were excluded because of insufficient data to confirm or refute the diagnosis. 61/71 remaining patients (85.9%) had a confirmed diagnosis of GPA. The PPV of this algorithm was 85.9% [75.6%–93.0%]. Adding positivity for cANCA to the algorithm increased the PPV to 94.4% [83.6%–99.3%].

Algorithms for identifying eosinophilic granulomatosis with polyangiitis

Of the 125 charts reviewed per algorithm-deriving healthcare system, 86 patients from System 1 and 117 patients from System 2 had sufficient data in their medical records to confirm or refute the diagnosis of EGPA and were included in the analysis. 19/86 (22.1%) patients had a confirmed diagnosis of EGPA at System 1 and 13/117 (11.1%) patients at System 2 based on the ACR criteria or the revised ICHCC definitions. Figure 2 shows the PPV and sensitivity of the 16 constructed case-finding algorithms for EGPA. The PPV and sensitivity of the algorithms in each healthcare system are detailed in Supplementary Table 3. Two algorithms: i) one including patients with eosinophilia *or* asthma and including the physician specialty (PPV: 90.0% at System 1 and 45.8% at System 2) and ii) another including patients with eosinophilia *or* asthma and including the physician specialty and an immunosuppressive medication (PPV: 90.0% at System 1 and 45.8% at System 2) had the highest average positive. Adding positivity for pANCA to the algorithms did not significantly increase the PPV and decreased the sensitivity.

58 out of 2,151,960 patients were identified at System 3 using the algorithm with the highest average PPV and the lowest variable numbers listed above (Patients with ICD9 code 446.4 including patients with eosinophilia *or* asthma and the physician specialty). 10 patients were excluded because of insufficient data to confirm or refute the diagnosis. 20/48 remaining patients (41.7%) had a confirmed diagnosis of EGPA. The PPV of this algorithm was: 41.7% [27.6%–56.7%] indicating a low PPV. However, adding positivity for pANCA to the algorithm increased the PPV to 100.0% [66.3%–100.0%].

Given the observed low PPV for the algorithm without the ANCA pattern at System 3, algorithms were reconstructed by including patients with both eosinophilia *and* asthma

instead of only one of these disease manifestations. Supplementary Table 4 lists the recalculated PPV and sensitivity of the algorithms for EGPA including patients with eosinophilia *and* asthma. Several algorithms with and without positivity for pANCA had a high average PPV.

21 patients were identified at healthcare system 3 using an algorithm that identified patients with ICD9 code 446.4 including patients with eosinophilia *and* asthma and the physician specialty. 18/21 had a confirmed diagnosis of EGPA (85.7%). The PPV of this algorithm was 85.7% [63.6%–93.9%]. Adding positivity for pANCA to the algorithm increased the PPV to 100% [15.8%–100%].

Algorithms for identifying microscopic polyangiitis

Of the 125 charts reviewed per algorithm-deriving healthcare system, 84 patients from System 1 and 116 patients from System 2 had sufficient data in their medical records to confirm or refute the diagnosis of MPA and were included in the analysis. 27/84 (32.1%) patients had a confirmed diagnosis of MPA at System 1 and 21/116 (18.1%) patients at System 2 based on the ACR criteria or the revised ICHCC definitions. Figure 3 shows the PPV and sensitivity of the 16 constructed case-finding algorithms for MPA. The PPV and sensitivity of the algorithms in each healthcare system are detailed in Supplementary Table 5. An algorithm including patients with one of the following diagnoses: 770.3 (Alveolar hemorrhage), 516.34 (Interstitial lung disease), 583.X (Glomerulonephritis), 584.X (Acute Kidney Failure), or 585.X (Chronic Kidney Disease), the encounter type, the physician specialty, and the immunosuppressive medications had the highest average positive predictive value (PPV: 85.7% at System 1 and 66.7% at System 2). Adding positivity for pANCA significantly increased the PPV across all algorithms (up to 100%) but decreased sensitivity.

36 out of 2,151,960 were identified at System 3 using the algorithm with the highest average PPV listed above. 5 patients were excluded because of insufficient data to confirm or refute the diagnosis. 19/31 (61.2%) had a confirmed diagnosis of MPA. The PPV of this algorithm was 61.2% [42.2%–78.2%]. Adding positivity for pANCA increased the PPV of this algorithm to 81.2% [75.6%–93.0%].

DISCUSSION

This study demonstrates that case-finding algorithms can accurately identify patients with GPA, EGPA, and MPA in large administrative databases. The successful algorithms, also known as “computable phenotypes,” contain various combinations of claims codes and laboratory variables and can be easily run within large administrative databases. Furthermore, the study underlines the inaccuracy of using ICD codes alone for the identification of patients with a specific type of AAV.

The ICD version 9 was initially published by the World Health Organization in 1978 (11). ICD9 codes often lack specificity and are subject to several errors resulting in low accuracy to identify patients with a specific disease (6, 11). For example, the numbers of patients having at least one ICD9 code for GPA, EGPA, and MPA in each healthcare system were

significantly higher than the reported prevalence of AAV (2, 12, 13) indicating a large number of incorrectly applied diagnoses in these populations. Additionally, the percentage of patients with GPA, EGPA, and MPA among patients carrying the corresponding ICD9 code was low (46.4% and 67.3%) for GPA, 22.1% and 21.8% for EGPA, and 32.1% and 18.1% for MPA at system 1 and system 2 respectively). Thus, using only one ICD9 code to identify patients with a specific type of AAV is subject to inaccuracy and may result in a high number of false positives.

Previous studies have shown that having a specific ICD9 code submitted twice or more, or one inpatient ICD9 code increases the PPV of case-finding algorithms (14–16). The same result was demonstrated in this study. Adding other variables such as the specialty of the physician involved in the care, the medications used specifically for the treatment of that condition, or a particular specific laboratory value also often increases the PPV of case-finding algorithms (17). In this study, the highest average PPVs were observed when the physician specialty (Rheumatologist, Allergist/Immunologist, Nephrologist, Pulmonologist, or Otorhinolaryngologist) was added to the algorithm. Adding a positive test for ANCA to the algorithms further improved the overall PPV for GPA, EGPA, and MPA in the validating system. Adding ANCA testing data was particularly important for the algorithm for MPA for which the PPV increased from 61.2% to 81.2% in the validating system, making results of pANCA testing an essential variable of this algorithm. Finally, including patients with asthma *or* eosinophilia to the algorithms for EGPA resulted in low PPV. Interestingly, many patients with GPA carried the diagnosis of asthma and to a less extent eosinophilia. This non-specific coding for asthma resulted in low PPVs for the algorithms for EGPA. However, including patients with both asthma *and* eosinophilia increased the PPV for the algorithms for EGPA but decreased the sensitivity (supplementary tables 3 and 4).

When utilizing a case-finding algorithm it is important that both PPV and sensitivity be considered, depending on the goal intended for the algorithm. Adding variables to a case-finding algorithm may increase the algorithm's PPV but often at the expense of reducing the algorithm's sensitivity (Figures 1, 2, and 3). While this study opted to validate algorithms with the highest PPV, case-finding algorithms with suboptimal PPV but high sensitivity may be the choice for some purposes.

This study has several important strengths. The data elements used to construct the draft algorithms incorporated items commonly used in routine clinical practice and for which there is little variability for coding into electronic health records. A stringent protocol was followed to identify and confirm cases in two healthcare systems to develop well-performing algorithms. The most promising algorithms were then validated in a third system to identify and confirm cases. Additionally, the complexities of overlapping clinical manifestations among the disease types (asthma and eosinophilia) were recognized and addressed to arrive at useful final results. Overall, the number of patients with these rare diseases studied in this project was similar to sample sizes used in similar projects involving common illnesses (20–22). Finally, the demonstration that these algorithms work well in three separate, large healthcare systems adds to the confidence in the generalizability of these findings.

This study also has a few limitations to consider. Some less specific ICD9 codes often used by physicians, such as “ANCA-associated vasculitis” or “arteritis” (447.8) were not examined in this project. However, while some patients might have been missed through this approach and thus lowering the sensitivity of the algorithms, the PPVs remain accurate and strong. The more recently introduced ICD10 codes were not studied. However data existing in ICD10 were insufficient to make definite conclusions due to its recent implementation. While ICD10 provides more granularity and specificity compared to ICD9, the diagnostic accuracy of ICD10 to identify patients with a clinical condition remains surprisingly low (7, 19).

In summary, case-finding algorithms can accurately identify patients with GPA, EGPA, and MPA using large administrative databases. These case-finding algorithms can be used to assemble population-based cohorts and facilitate future research in healthcare economics, comparative effectiveness, drug safety, and disease outcomes in patients with AAV.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: The Vasculitis Patient-Powered Research Network is a participant of PCORnetSM, the National Patient-Centered Clinical Research Network, an initiative funded by the Patient-Centered Outcomes Research Institute (PCORI). The Vasculitis Patient-Powered Research Network’s participation was funded through PCORI Awards (PPRN-1306-04758 and 1IP2PI000603)

Dataset from the second healthcare system were obtained from Vanderbilt University Medical Center’s Synthetic Derivative which is supported by institutional funding and by the Vanderbilt CTSA grant ULTR000445 from NCATS/NIH

Dataset from the third healthcare system were obtained from the University of Kansas Medical Center Healthcare Enterprise Repository for Ontological Narration (HERON) clinical data repository which is supported by institutional funding and by the University of Kansas Medical Center CTSA grant UL1TR000001 from NCRR/NIH

References

1. Lane SE, Watts RA, Shepstone L, Scott DG. Primary systemic vasculitis: clinical features and mortality. *QJM : monthly journal of the Association of Physicians*. 2005; 98(2):97–111. [PubMed: 15655098]
2. Watts RA, Mahr A, Mohammad AJ, Gatenby P, Basu N, Flores-Suarez LF. Classification, epidemiology and clinical subgrouping of antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. 2015; 30(Suppl 1):i14–22.
3. Raimundo K, Farr AM, Kim G, Duna G. Clinical and Economic Burden of Antineutrophil Cytoplasmic Antibody-associated Vasculitis in the United States. *The Journal of rheumatology*. 2015; 42(12):2383–91. [PubMed: 26523026]
4. Gibson A, Stamp LK, Chapman PT, O'Donnell JL. The epidemiology of Wegener’s granulomatosis and microscopic polyangiitis in a Southern Hemisphere region. *Rheumatology*. 2006; 45(5):624–8. [PubMed: 16368729]
5. Boyd M, Specks U, Finkelstein JD. Accuracy of the ICD-9 code for identification of patients with Wegener’s granulomatosis. *The Journal of rheumatology*. 2010; 37(2):474. [PubMed: 20147488]

6. Myers RP, Leung Y, Shaheen AA, Li B. Validation of ICD-9-CM/ICD-10 coding algorithms for the identification of patients with acetaminophen overdose and hepatotoxicity using administrative data. *BMC health services research*. 2007; 7:159. [PubMed: 17910762]
7. Quan H, Li B, Saunders LD, Parsons GA, Nilsson CI, Alibhai A, et al. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. *Health services research*. 2008; 43(4):1424–41. [PubMed: 18756617]
8. Jennette JC, Falk RJ, Bacon PA, Basu N, Cid MC, Ferrario F, et al. 2012 revised International Chapel Hill Consensus Conference Nomenclature of Vasculitides. *Arthritis and rheumatism*. 2013; 65(1):1–11. [PubMed: 23045170]
9. Harrold LR, Andrade SE, Eisner M, Buist AS, Go A, Vollmer WM, et al. Identification of patients with Churg-Strauss syndrome (CSS) using automated data. *Pharmacoepidemiology and drug safety*. 2004; 13(10):661–7. [PubMed: 15386588]
10. Fries JF, Hunder GG, Bloch DA, Michel BA, Arend WP, Calabrese LH, et al. The American College of Rheumatology 1990 criteria for the classification of vasculitis. Summary. *Arthritis and rheumatism*. 1990; 33(8):1135–6. [PubMed: 2202312]
11. O'Malley KJ, Cook KF, Price MD, Wildes KR, Hurdle JF, Ashton CM. Measuring diagnoses: ICD code accuracy. *Health services research*. 2005; 40(5 Pt 2):1620–39. [PubMed: 16178999]
12. Ntatsaki E, Watts RA, Scott DG. Epidemiology of ANCA-associated vasculitis. *Rheumatic diseases clinics of North America*. 2010; 36(3):447–61. [PubMed: 20688243]
13. Scott DG, Watts RA. Epidemiology and clinical features of systemic vasculitis. *Clinical and experimental nephrology*. 2013; 17(5):607–10. [PubMed: 23843034]
14. Ng B, Aslam F, Petersen NJ, Yu HJ, Suarez-Almazor ME. Identification of rheumatoid arthritis patients using an administrative database: a Veterans Affairs study. *Arthritis care & research*. 2012; 64(10):1490–6. [PubMed: 22623324]
15. Goldberg D, Lewis J, Halpern S, Weiner M, Lo Re V 3rd. Validation of three coding algorithms to identify patients with end-stage liver disease in an administrative database. *Pharmacoepidemiology and drug safety*. 2012; 21(7):765–9. [PubMed: 22674685]
16. Moores KG, Sathe NA. A systematic review of validated methods for identifying systemic lupus erythematosus (SLE) using administrative or claims data. *Vaccine*. 2013; 31(Suppl 10):K62–73. [PubMed: 24331075]
17. Chung CP, Rohan P, Krishnaswami S, McPheeters ML. A systematic review of validated methods for identifying patients with rheumatoid arthritis using administrative or claims data. *Vaccine*. 2013; 31(Suppl 10):K41–61. [PubMed: 24331074]
18. Outland B, Newman MM, William MJ. Health Policy Basics: Implementation of the International Classification of Disease, 10th Revision. *Annals of internal medicine*. 2015; 163(7):554–6. [PubMed: 26390305]
19. De Coster C, Li B, Quan H. Comparison and validity of procedures coded With ICD-9-CM and ICD-10-CA/CCI. *Medical care*. 2008; 46(6):627–34. [PubMed: 18520318]
20. Kim SY, Servi A, Polinski JM, Mogun H, Weinblatt ME, Katz JN, et al. Validation of rheumatoid arthritis diagnoses in health care utilization data. *Arthritis research & therapy*. 2011; 13(1):R32. [PubMed: 21345216]
21. Andrade SE, Scott PE, Davis RL, Li DK, Getahun D, Cheetham TC, et al. Validity of health plan and birth certificate data for pregnancy research. *Pharmacoepidemiology and drug safety*. 2013; 22(1):7–15. [PubMed: 22753079]
22. Katz JN, Barrett J, Liang MH, Bacon AM, Kaplan H, Kieval RI, et al. Sensitivity and positive predictive value of Medicare Part B physician claims for rheumatologic diagnoses and procedures. *Arthritis and rheumatism*. 1997; 40(9):1594–600. [PubMed: 9324013]

Figure 1.

Average positive predictive (PPV) and sensitivity of algorithms for granulomatosis with polyangiitis (GPA, Wegener's) at healthcare systems 1 and 2. Algorithms are arranged by the smallest to highest PPV. Highlighted on the right is the algorithm with the highest average PPV not including the ANCA pattern.

PPV: Positive predictive value.

ICD9: ICD9 code 446.4 excluding patients with ICD9 codes 283.3 (eosinophilia) or 493.X (asthma).

Encounter: 1 inpatient ICD9 code 446.4 on 3 consecutive days or 2 ICD9 codes 446.4 3 months apart.

Specialty: Rheumatologist, Allergist/Immunologist, Nephrologist, Pulmonologist, or Otorhinolaryngologist involved in the care of the patient.

Medications: Any of the medications listed in Supplementary Table 6.

cANCA: Positive test for cytoplasmic pattern of anti-neutrophil cytoplasmic antibody

Figure 2.

Average positive predictive value (PPV) and sensitivity of algorithms for eosinophilic granulomatosis with polyangiitis (EGPA, Churg-Strauss) in healthcare systems 1 and 2 (including patients with eosinophilia *OR* asthma). Algorithms are arranged by the smallest to highest PPV. Highlighted on the right are the algorithms with the highest average PPV not including the ANCA pattern (both algorithms had the same average PPV and sensitivity).

PPV: Positive predictive value.

EGPA: eosinophilic granulomatosis with polyangiitis (Churg-Strauss).

ICD9: ICD9 code 446.4 including patients with ICD9 codes 283.3 (eosinophilia) *or* 493.X (asthma).

Encounter: 1 inpatient ICD9 code 446.4 on 3 consecutive days or 2 ICD9 codes 446.4 3 months apart.

Specialty: Rheumatologist, Allergist/Immunologist, Nephrologist, Pulmonologist, or Otorhinolaryngologist involved in the care of the patient.

Medications: Any of the medications listed in Supplementary Table 6.

pANCA: Positive test for perinuclear pattern of anti-neutrophil cytoplasmic antibody.

Figure 3.

Average positive predictive value (PPV) and sensitivity of algorithms for MPA in healthcare systems 1 and 2. Algorithms are arranged by the smallest to highest PPV. Highlighted on the right is the algorithm with the highest average PPV not including the ANCA pattern.

PPV: Positive predictive value.

ICD9: ICD9 code 446.4 ICD9 code 446.4 including patients with one of the following diagnoses: 770.3 (Alveolar hemorrhage), 516.34 (Interstitial lung disease), 583.X (Glomerulonephritis), 584.X (Acute Kidney Failure), or 585.X (Chronic Kidney Disease)

Encounter: 1 inpatient ICD9 code 446.4 on 3 consecutive days or 2 ICD9 codes 446.4 3 months apart.

Specialty: Rheumatologist, Allergist/Immunologist, Nephrologist, Pulmonologist, or Otorhinolaryngologist involved in the care of the patient.

Medications: Any of the medications listed in Supplementary Table 6.

pANCA: Positive test for perinuclear pattern of anti-neutrophil cytoplasmic antibody

Table 1

Total number of patients and number of patients with at least one ICD9 code (446.4 and 446.0) in Healthcare Systems 1 and 2

	Healthcare System 1	Healthcare System 2
Number of patients in the healthcare system	3,161,862	2,443,098
Number of patients with at least one ICD9 code 446.4 (GPA/EGPA)	486	559
Number of patients with at least one ICD9 code 446.0 (MPA)	221	385

ICD9: International Classification of Diseases version 9; GPA: Granulomatosis with polyangiitis (Wegener's); EGPA: Eosinophilic granulomatosis with polyangiitis (Churg-Strauss); MPA: Microscopic polyangiitis. 125 charts per diagnosis per algorithm-deriving healthcare system were randomly-selected from the number of patients with at least one ICD9 code 446.4 or 446.0.