

Effective Health Care Research Reports

Number 7

Computerized Definitions Showed High Positive Predictive Values for Identifying Reasons for Hospitalizations in a Medicaid Population With Rheumatoid Arthritis

Carlos G. Grijalva, M.D., M.P.H.
Cecilia P. Chung, M.D., M.P.H.
C. Michael Stein, M.D.
Patricia S. Gideon, R.N.
Shannon M. Dyer
Edward F. Mitchel, Jr., M.S.
Marie R. Griffin, M.D., M.P.H.

Research from the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) Network



Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

October 2008

The DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) network is part of AHRQ's Effective Health Care Program. It is a collaborative network of research centers that support the rapid development of new scientific information and analytic tools. The DEcIDE network assists health care providers, patients, and policymakers seeking unbiased information about the outcomes, clinical effectiveness, safety, and appropriateness of health care items and services, particularly prescription medications and medical devices.

This report is based on research conducted by the Vanderbilt University DEcIDE (Developing Evidence to Inform Decisions about Effectiveness) Center under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. HSA290200500421). The AHRQ Task Order Officer for this project was Gurveet Randhawa, M.D., M.P.H.

The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

None of the authors has a financial interest in any of the products discussed in this report.

This report has also been published in edited form: Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF Jr, Griffin MR. Computerized definitions showed high positive predictive values for identifying hospitalizations for congestive heart failure and selected infections in Medicaid enrollees with rheumatoid arthritis. *Pharmacoepidemiol Drug Saf* 2008 Jun 10 [Epub].

Suggested Citation:

Grijalva CG, Chung CP, Stein CM, Gideon PS, Dyer SM, Mitchel EF Jr, Griffin MR. Computerized definitions showed high positive predictive values for identifying reasons for hospitalizations in a Medicaid population with rheumatoid arthritis. *Effective Health Care Research Report No. 7.* (Prepared by Vanderbilt University DEcIDE Center Under Contract No. HSA290200500421.) Rockville, MD: Agency for Healthcare Research and Quality. October 2008. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Contents

Introduction.....	1
Methods.....	1
Results.....	3
Discussion.....	4
Conclusion.....	5
Acknowledgments.....	6
References.....	7
Appendix. ICD9-CM Codes Used in Computerized Definitions.....	8

Author affiliations:

Carlos G. Grijalva, M.D., M.P.H.¹

Cecilia P. Chung, M.D., M.P.H.²

C. Michael Stein, M.D.²

Patricia S. Gideon, R.N.¹

Shannon M. Dyer¹

Edward F. Mitchel, Jr., M.S.¹

Marie R. Griffin, M.D., M.P.H.^{1,2,3}

¹Department of Preventive Medicine, Vanderbilt University School of Medicine, Nashville, TN.

²Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN.

³Mid-South Geriatric Research Education and Clinical Center and Clinical Research Center of Excellence, VA Tennessee Valley Healthcare System, Nashville, TN.

Abstract

Purpose: Computerized definitions are used to identify serious infections and congestive heart failure leading to hospitalizations in studies of medication safety. However, information on their accuracy is limited. We evaluated the ability of computerized definitions to identify these conditions among patients diagnosed with rheumatoid arthritis.

Methods: Medical charts were randomly selected from a systematic sample of hospitalizations identified in a cohort of Medicaid patients with rheumatoid arthritis. We calculated positive predictive values (PPV) for computerized definitions for community-acquired pneumonia, invasive pneumococcal disease, sepsis, opportunistic mycoses and congestive heart failure using charts reviews as gold standard and computed inter-reviewer agreement statistics.

Results: From 2667 hospitalizations, 336 (13%) records were selected for review. A total of 277 charts (83%) were available. Based on any discharge diagnosis, PPVs for hospitalizations due to community-acquired pneumonia, invasive pneumococcal disease, sepsis and opportunistic mycoses were 84%, 100%, 80% and 62%, respectively. Restricting definitions to primary diagnoses yielded higher PPVs, 95% for pneumonia and 100% for other diagnoses. The PPV of a primary diagnosis for congestive heart failure was 100%. Inter-reviewer agreement was at least 77% for any outcome.

Conclusion: These findings suggest that computerized definitions can identify common conditions leading to hospitalization in Medicaid patients with rheumatoid arthritis

Introduction

Administrative databases are commonly used in epidemiologic assessments of medication safety. These resources are attractive because the information recorded represents large populations in the setting of routine healthcare utilization and data are collected systematically, free of recall bias and without regard to medication exposure or disease status. Discharge diagnosis codes, collected for administrative purposes, are frequently used to create computerized definitions for study outcomes or covariates.¹ Although these codes are commonly used in research, information on their actual ability to identify specific outcomes is limited.²

Inaccurate computerized definitions would result in misclassification of outcomes or covariates and such misclassification makes it more difficult to detect true associations if they exist. Although time and resource consuming, evaluating the accuracy of outcome identification is often necessary to assure the validity of epidemiologic studies.

Tennessee Medicaid (TennCare) databases have been used extensively for post marketing studies of medication safety and we are evaluating the safety of tumor necrosis factor alpha (TNF- α) antagonists in a cohort of patients with rheumatoid arthritis (RA). Although these TNF- α antagonists offer benefits to patients with RA, exposure to these drugs has been associated with adverse events such as serious infections and congestive heart failure.³⁻⁶ Available clinical trial data are limited and insufficient to address these safety concerns conclusively, thus epidemiologic studies performed in large databases can contribute to the assessment of the safety of new therapies for RA.

Establishing measurements of the accuracy of computerized definitions for selected clinical conditions will facilitate studies on the safety of TNF- α antagonists and other medications. In this study, we determined the positive predictive value (PPV) of computerized case definitions for selected common (pneumonia and congestive heart failure) and less frequent (sepsis, invasive pneumococcal diseases and opportunistic mycoses) causes of hospitalization in patients with RA.

Methods

TennCare, the state-based capitated model program in Tennessee, covers those who are Medicaid-eligible and those who are uninsured or lack other access to health care. From 1995 through 2004, we identified TennCare enrollees aged ≥ 18 years diagnosed with RA (Excluding codes for unspecified inflammatory arthropathies, Appendix). Patients entered the cohort when they met one of three selection criteria: A hospitalization discharge diagnosis of RA; one ambulatory visit that resulted in a diagnosis of RA and a prescription for a disease modifying anti-rheumatic drug (DMARD); or two ambulatory visits (at least 1 month apart) that resulted in a diagnosis of RA. This study was approved by the Vanderbilt Institutional Review Board and by the Bureau of TennCare.

Cohort members had at least 365 days of continuous enrollment before entering the cohort to allow the collection of baseline characteristics. To limit potential sources of poor follow-up we excluded patients with solid organ transplantation, HIV/AIDS, cancer -except non-melanoma skin cancers-, and serious renal (patients on hemodialysis), liver (hepatic coma and hepatorenal syndrome), or respiratory diseases (acute or chronic respiratory failure) identified during the

baseline period. Cohort members were required to have at least one prescription filled during baseline to assure access to medication benefits. Follow-up started on the date when selection criteria were met and continued through the end of the study (12/31/2004), date of death, diagnosis of a serious medical condition or loss of TennCare enrollment, whichever came earliest.

Computerized definitions for community-acquired pneumonia, sepsis, congestive heart failure, invasive pneumococcal diseases and opportunistic mycoses that resulted in hospitalization were created using ICD9-CM discharge diagnoses (Appendix) keeping track of the diagnosis field position (i.e. principal or secondary). We restricted our study to hospitalization records, because outpatient records were not available for review. Since the occurrence of nosocomial events might depend on factors not typically recorded in administrative databases, we identified events that resulted in hospitalization rather than nosocomial events.⁷

A computer-generated random number was assigned to each hospitalization record identified and to each eligible hospital. For logistical reasons we restricted our chart review to hospitals located within a radius of 200 miles around Nashville, TN. Hospitalization records were sorted by their random numbers and the numbers allocated to their respective hospitals. Then, records for review were selected according to this sorted list. We maximized the number of sampled hospitals, requiring a minimum of 5 and no more than 20 charts per hospital for review.

Trained nurses traveled throughout the State of Tennessee, visited selected hospitals and retrieved medical records for review. The information was collected using a pilot-tested abstraction form for each condition of interest. Two study investigators (CGG, CPC), blinded to the medication exposure status reviewed the abstraction forms separately. Inter-reviewer agreement (i.e. kappa) statistic was computed and a third reviewer (MRG), who was also blinded to exposure status, resolved all disagreements.

Using the medical chart reviews as gold standard, we calculated the PPV of our definitions, that is, the proportion of cases identified by our computerized definitions that were subsequently confirmed through chart review. Episodes of community-acquired pneumonia were deemed as confirmed if, on admission, pneumonia was considered the main reason for hospitalization; and, at the end of the hospitalization, pneumonia was considered the main disease present at admission by the treating physician. Our pneumonia abstraction form was a modified version of a previously validated tool^{8,9} and similar forms were designed for other study outcomes. Pneumonia that developed during the course of the hospitalization (i.e. nosocomial), was considered not confirmed.

An episode of sepsis was confirmed if it was considered the main reason for hospitalization and the diagnosis was accompanied by the isolation/identification of pathogenic microorganisms from normally sterile tissue, fluid or body cavity.¹⁰ Sepsis that developed during the course of a hospitalization due to other reasons was considered nosocomial. Similarly, the confirmation of hospitalization for invasive pneumococcal disease required the identification or isolation of *Streptococcus pneumoniae* from normally sterile tissue, fluid or body cavity.

We aimed to identify episodes of new onset or exacerbations of congestive heart failure. Unlike other study outcomes, congestive heart failure commonly has a chronic course and for some patients a hospital discharge diagnosis represents a comorbidity rather than the main cause of hospitalization. During the pilot testing of our abstraction form we reviewed 45 medical charts randomly selected from a local hospital. Seven of 19 charts for congestive heart failure (identified by codes listed in any diagnosis field) had congestive heart failure as the main reason

for admission, and all these confirmed cases had congestive heart failure listed as the primary diagnosis. Unconfirmed cases had congestive heart failure as a comorbidity but were admitted for different medical conditions. Therefore, we restricted our evaluation of this outcome to primary diagnoses. An episode of congestive heart failure was deemed as confirmed if congestive heart failure was considered the main reason for hospitalization and at the end of the hospitalization, treating physicians considered it to be the main cause of the hospitalization.

Episodes of opportunistic mycoses were deemed as confirmed if they were considered the main reason for hospitalization and they were accompanied by the identification/isolation of specific organisms from tissue, fluid or body cavity or by the presence of other compatible laboratory results.¹¹

Results

Within the study period, 14932 patients diagnosed with RA met our selection criteria. The median age of cohort members was 55 years; most patients were female (77%), white (73%), and lived in standard metropolitan statistical areas (53%). At the beginning of follow-up, 591 (4%) subjects were nursing home residents and 49% had disability specified as their qualifying condition for Medicaid enrollment.

Using our computerized definitions, we identified 2667 hospitalizations (including 2427 for pneumonia or congestive heart failure and 240 for sepsis, invasive pneumococcal disease or opportunistic mycoses). A total of 336 (13%) hospitalizations distributed in 17 Tennessee hospitals were identified for review. Among these hospitalizations, 59 (18%) had no medical chart available. We reviewed 161 charts for pneumonia, 38 for congestive heart failure (3 with a secondary pneumonia diagnosis), 45 for sepsis and 7 for invasive pneumococcal disease. In addition, we reviewed 26 charts for opportunistic mycoses including candidiasis (n=21), cryptococcosis (2), aspergillosis (2) and histoplasmosis (1).

Blinded adjudication classified 135 of 161 cases that met our computerized definition of community-acquired pneumonia, as confirmed, yielding a PPV of 84% (95% CI: 77-89) for pneumonia codes listed in any diagnosis field. Among the confirmed cases, 130 (96%) had radiological reports with compatible findings, 3 did not show acute changes and 2 did not have radiological reports available. Twenty-six (16%) of 161 eligible cases were not confirmed (Table 1) and the inter-reviewer agreement was 91% (kappa=0.6213, p<0.001). Of the 26 cases that were not confirmed, 13 (50%) represented nosocomial pneumonias, 4 (15%) were hospitalized for chronic obstructive pulmonary disease/asthma, 5 (19%) abdominal pain, 2 (8%) fractures, 1 (4%) gastro-intestinal bleeding and 1 (4%) neoplasia. The PPV for pneumonia codes listed in the primary diagnosis field was 95% (95% CI: 90-98).

The overall PPV for sepsis was 80% (95% CI: 65-90) and the agreement between reviewers was 82% (kappa=0.2437, p=0.0387). Among confirmed cases, 34 (94%) had positive blood cultures and 2 had positive cultures from other tissues. When the analysis was restricted to primary diagnosis fields, the PPV was 100%, although the number of records reviewed was small (Table 1). The PPV for codes in secondary fields was 75% (95% CI: 58-88). Among the 9 sepsis episodes that were considered not confirmed, 5 (56%) were nosocomial cases, 2 (22%) considered the isolated agent as a contaminant, 1 (11%) lacked a positive isolation of a microorganism and 1 (11%) was an episode of gastro-intestinal bleeding.

Seven medical charts for invasive pneumococcal disease (3 in primary and 4 in secondary diagnosis fields) were identified (1 case of meningitis and 6 of bacteremia). All cases were

confirmed (PPV: 100%) and had positive cultures (6 blood and 1 cerebrospinal fluid). The inter-reviewer agreement was 100% ($\kappa=1$, $p=0.0041$). All records identified by codes in secondary fields had a primary discharge diagnosis of pneumococcal pneumonia.

All episodes of congestive heart failure based on the primary discharge diagnosis ($n=38$) were confirmed yielding an estimated PPV of 100% (Table 1). The inter-reviewer agreement was 95% ($\kappa=0.7295$, $p<0.001$). Among confirmed cases, 23 (61%) records had an echocardiography compatible with congestive heart failure, 13 (34%) did not have an echocardiography performed, 1 (3%) had a report indicating an ejection fraction of 65% without further details and 1 had a report indicating normal ventricular function in a patient with a history of multiple coronary bypass procedures (the latter two patients had chest x-rays compatible with congestive heart failure on admission). All three cases with concurrent codes for congestive heart failure (primary diagnosis) and pneumonia (secondary diagnosis) were confirmed.

Finally, we confirmed 16 of 26 records for opportunistic mycoses, yielding a PPV of 62% (95% CI: 41-80). The agreement between reviewers was 77% ($\kappa=0.4507$, $p=0.003$). Codes listed in primary and secondary fields yielded PPVs of 100% and 50%, respectively, but the number of medical charts reviewed was small (Table 1). We confirmed 13 out of 21 records for opportunistic candidiasis, including 3 disseminated candidiasis, 3 candidiasis of the lung and 7 esophageal candidiasis (PPV: 62%, 95% CI: 38-82). Among confirmed candidiasis cases, 8 had positive cultures and 5 (esophageal candidiasis) were diagnosed during endoscopy and treated as such. In addition, we reviewed and confirmed 2 records for cryptococcosis (1 case of meningitis and 1 of pneumonia) and 1 for aspergillosis of the lung.

Discussion

Our findings indicate that computerized definitions can identify common conditions leading to hospitalization in patients with RA and could be used in assessments of medication safety. Highest PPVs were consistently obtained from definitions based on primary discharge diagnoses and the inter-reviewer agreement was high.

Computerized definitions for community-acquired pneumonia, sepsis and congestive heart failure, based on primary diagnoses among patients with RA, had high PPVs (greater than 95%). In contrast, codes listed in other diagnostic fields were less accurate and often represented nosocomial conditions. Similar studies conducted in the general population and using other sources of data have reported PPVs ranging from 89% through 98% for bacteremia/septicemia codes listed in any diagnosis field and including nosocomial cases.^{12, 13} Similarly, a study that used primary discharge diagnoses from a regional Veterans Affairs system showed that computerized definitions for bacteremia had PPVs ranging from 86%-91%.⁷

Assessing the ability of our definitions to identify rare conditions, such as invasive pneumococcal diseases or opportunistic mycoses was challenging because of the limited number of records available. A recent study reported high PPVs for cryptococcosis diagnosis codes and low PPVs for systemic candidiasis diagnosis codes among hospitalized veterans. However, the number of records reviewed was small. Further research would help clarify the accuracy of these definitions. Meanwhile, systematic confirmation of these rare outcomes would be necessary.

The interpretation of our findings requires the consideration of several caveats. Although review of medical charts is commonly considered the gold standard for validation of computerized definitions, there are limitations to information recorded in the charts. Previous studies have shown that procedures and diagnoses can be left undocumented in medical charts.²

Although we used a previously validated definition for pneumonia, that definition as well as other definitions for this study was based in large part on the clinical determination of the treating provider(s). Nevertheless, using the available information and without regard to gold standard determinations, the agreement between independent reviewers was high.

In our study, it was not possible to estimate the sensitivity or the specificity of our definitions. Such determination would require review of all or at least a random sample of medical records without regard to presence or absence of the disease of interest, that is, medical records in which the computerized definitions missed the actual reason for hospitalization (false negatives) and those records that were correctly identified as non cases (true negatives). Instead, we estimated the proportion of cases that were correctly identified by our computerized definitions (PPV). This PPV depends on disease prevalence and the specificity of the definitions⁷, and when disease prevalence is low, a high PPV indicates high specificity, allowing an unbiased estimation of relative risks, regardless of the sensitivity of the outcome definitions^{7, 14}, as long as the sensitivity is similar for all comparison groups.

Although TennCare enrollees are not representative of the general population, they share the same hospitals as other Tennesseans and it is likely that coding for these patients is similar to the rest of the State population. Nevertheless, although our study was restricted to patients fulfilling selection criteria for RA, the comparison of our results with results obtained from other non-RA populations requires some additional considerations. For instance, whether or not the use of RA-related medications (i.e. immunosuppressives) may affect the clinical manifestations of infection and result in differences in disease identification and coding among patients with RA, is unknown.¹⁵

Conclusion

The increasing availability of large databases provides new opportunities for research. However, the design of strategies to identify study outcomes and covariates requires careful attention and information on the accuracy of these strategies is necessary. Our study demonstrates that computerized definitions can identify selected common conditions leading to hospitalization among Medicaid patients diagnosed with RA, so that these definitions can be used for evaluations of medications safety.

Table 1. Positive predictive value (PPV) of computerized definitions for events leading to hospitalization in TennCare patients diagnosed with rheumatoid arthritis

Computerized Definition	Reviewed	Confirmed	PPV (%)	95% CI
Community-acquired pneumonia (any field)	161	135	84	(77-89)
Primary	108	103	95	(90-98)
Secondary only	53	32	60	(46-74)
Sepsis (any field)	45	36	80	(65-90)
Primary	9	9	100*	-
Secondary only	36	27	75	(58-88)
Invasive Pneumococcal disease (any field)	7	7	100*	-
Primary	3	3	100*	-
Secondary only	4	4	100*	-
Congestive Heart failure (primary)	38	38	100*	-
Opportunistic mycoses (any field)	26	16	62	(41-80)
Primary	6	6	100*	-
Secondary only	20	10	50	(27-73)

*All records were confirmed

Acknowledgments

We are indebted to the Tennessee Bureau of TennCare of the Department of Finance and Administration, which provided the data. We are also indebted to Dr. Lisa Jackson for facilitating abstraction forms used to design the validation tools for this study.

References

1. Strom BL. Data validity issues in using claims data. *Pharmacoepidemiol Drug Saf* 2001;10:389-92.
2. Wilchesky M, Tamblyn RM, Huang A. Validation of diagnostic codes within medical services claims. *J Clin Epidemiol* 2004;57:131-41.
3. Dixon WG, Watson K, Lunt M, Hyrich KL, Silman AJ, Symmons DP. Rates of serious infection, including site-specific and bacterial intracellular infection, in rheumatoid arthritis patients receiving anti-tumor necrosis factor therapy: Results from the British Society for Rheumatology Biologics Register. *Arthritis Rheum* 2006;54:2368-76.
4. Curtis JR, Patkar N, Xie A, Martin C, Allison JJ, Saag M, et al. Risk of serious bacterial infections among rheumatoid arthritis patients exposed to tumor necrosis factor alpha antagonists. *Arthritis Rheum* 2007;56:1125-33.
5. Khanna D, McMahon M, Furst DE. Anti-tumor necrosis factor alpha therapy and heart failure: what have we learned and where do we go from here? *Arthritis Rheum* 2004;50:1040-50.
6. St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P, et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum* 2004;50:3432-43.
7. Schneeweiss S, Robicsek A, Scranton R, Zuckerman D, Solomon DH. Veteran's affairs hospital discharge databases coded serious bacterial infections accurately. *J Clin Epidemiol* 2007;60:397-409.
8. Jackson ML, Neuzil KM, Thompson WW, Shay DK, Yu O, Hanson CA, et al. The burden of community-acquired pneumonia in seniors: results of a population-based study. *Clin Infect Dis* 2004;39:1642-50.
9. Jackson LA, Neuzil KM, Yu O, Benson P, Barlow WE, Adams AL, et al. Effectiveness of pneumococcal polysaccharide vaccine in older adults. *N Engl J Med* 2003;348:1747-55.
10. Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, et al. 2001 SCCM/ESICM/ACCP/ATS/SIS International Sepsis Definitions Conference. *Crit Care Med* 2003;31:1250-6.
11. O'Shaughnessy EM, Shea YM, Witebsky FG. Laboratory diagnosis of invasive mycoses. *Infect Dis Clin North Am* 2003;17:135-58.
12. Martin GS, Mannino DM, Eaton S, Moss M. The epidemiology of sepsis in the United States from 1979 through 2000. *N Engl J Med* 2003;348:1546-54.
13. Hackam DG, Mamdani M, Li P, Redelmeier DA. Statins and sepsis in patients with cardiovascular disease: a population-based cohort analysis. *Lancet* 2006;367:413-8.
14. Green MS. Use of predictive value to adjust relative risk estimates biased by misclassification of outcome status. *Am J Epidemiol* 1983;117:98-105.
15. Curtis JR, Martin C, Saag KG, Patkar NM, Kramer J, Shatin D, et al. Confirmation of administrative claims-identified opportunistic infections and other serious potential adverse events associated with tumor necrosis factor alpha antagonists and disease-modifying antirheumatic drugs. *Arthritis Rheum* 2007;57:343-6.

Appendix. ICD9-CM Codes Used in Computerized Definitions

Outcome	ICD9-CM
Rheumatoid arthritis	714.0, 714.1, 714.2, 714.3, 714.30, 714.31, 714.32, 714.33, 714.4 and 714.81
Pneumonia	480*, 481*, 482*, 483*, 484*, 485*, 486* and 487.0
Congestive Heart failure	428*
Bacteremia/sepsis	003.1, 036.2, 785.52, 790.7 and 038*
Invasive pneumococcal disease	320.1, 038.2 and 567.1
Opportunistic mycoses	117.3, 518.6, 484.6, 112.4, 112.5, 112.81, 112.83, 112.84, 114*, 117.5, 321.0 and 115*

*Indicates all codes starting with this number