Validity of Identification Methods of Lower Extremity Amputation in the Veterans Health Administration Electronic Medical Records

Morgan Meadows

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2020

Committee:

Alyson Littman

Edward Boyko

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Validity of Identification Methods of Lower Extremity Amputation in the Veterans Health Administration Electronic Medical Records

Morgan Meadows

Chair of the Supervisory Committee:

Alyson Littman

Department of Epidemiology

It is not currently known what the best method to identify individuals with lower extremity amputation (LEA) is. The purpose of this study is to determine the positive predictive value (PPV) of algorithms used to identify patients with LEA using Veterans Health Administration (VHA) electronic medical records (EMR) and to determine if PPV varies by age, gender, and race. 685 patients identified as having at least one diagnosis or procedure code for LEA and being alive were mailed a survey and asked to provide self-reported LEA status. We received 441 (64%) responses. We calculated PPV estimates and false negative percentages for nine algorithms. Algorithm 1, allowing any procedure or diagnosis code for LEA, consistently had the lowest PPV estimate across both the entire sample and all subgroups. Algorithms requiring at least one procedure code or two or more diagnosis codes generally had high PPVs, with the algorithm with the highest PPV estimate varying among subgroups and for the entire sample. In general, increasing restriction in an algorithm resulted in a higher PPV and proportion of false negatives. Ultimately, the best choice of algorithm depends on specific study needs.

Introduction

It is estimated that over 1.6 million persons in the United States are living with an amputation, including approximately 90,000 Veterans seen at Veterans Health Administration (VHA) facilities in 2017.[20,22] Amputees require long term and specialized care in order to optimize their reintegration into society.[21] The number of persons living with an amputation is expected to double by 2050.[22] Approximately 65% of amputees have a lower extremity amputation (LEA), with over half of those being major amputations.[22] As the number of amputees grows, we must be able to accurately identify the population with a LEA in order to assess health outcomes, provide high quality care, and dedicate the appropriate amount of resources.

Electronic medical records (EMRs) can be a wealth of information for researchers. These data allow researchers to conduct large studies at low costs as well as to study rare conditions and specific subpopulations. However, there are limitations to using EMRs. Since EMRs include data collected primarily for clinical and billing use, there is often questionable accuracy, errors in coding, and lack of quality control.[8] There are billing incentives that often dictate coding, miscoding, and omissions that impact the accuracy of the codes. Given these limitations and the value of using EMRs for research, it is critical to validate algorithms to identify populations of interest.

Prior studies validating EMR codes against a gold-standard have shown variability in positive predictive values (PPV), which ranged from 27%-98%.[1,2,3,6,7,9,11,13,15,17,18] The variability in PPV indicates that the ability of codes to accurately identify the intended population depends on factors including diagnosis, quality control measures, reimbursement model, and potentially other factors. Conditions with the highest PPVs were procedures, such as knee replacement and hip replacement, and common conditions, such as asthma.[3,11] Algorithms for serious conditions, such as sickle cell disease, also gave high PPVs when part of the algorithm accounted for number of visits and hospital admission.[15] Those with lower PPVs included events that could vary in severity, such as nonsteroidal anti-inflammatory drug (NSAID) related upper gastrointestinal events, or conditions requiring multiple tests and/or repeated testing, such as Hepatitis B for which a diagnosis includes results from three separate tests and often a set of follow-up tests 6 months later.[3,17] Additionally, for conditions with varying severity, such as early onset dementia and PTSD, algorithms often yielded the highest PPVs for those with specific other comorbidities.[9,13] Individuals without specific comorbidities were often found to be misclassified. Furthermore, diagnoses and events that had a short list of potential codes, such as palliative care, had high PPVs.[6] For diagnoses and events with long lists of potential codes, such as diabetes, more complex algorithms were required to achieve high PPVs.[16] Options for gold standards include chart review, prescription records, and self-report.[1,2,3,6,7,16,17,18] Self-report is a reliable gold standard for diagnoses and procedures that are objective and easy to self-determine. LEA is a good example of when self-report can be used as the gold standard as it is easy for individuals without cognitive impairment to know whether their legs were amputated.

To more efficiently and effectively study Veterans with LEA, it is important to evaluate the PPV of different algorithms to determine which approach minimizes false positives (those who have codes for LEA but self-report not having LEA) and maximizes true positives. Understanding how PPV may vary by gender, age, and race is similarly important. Much of the research done on LEA has focused on health outcomes in Veterans where males predominate.[12,14] Given the increasing number of women Veterans and subsequently number of women Veterans with LEA, it is important to examine how PPV may vary by gender.[10] Furthermore, the distribution of amputation etiology differs with age, with traumatic amputation more prevalent in younger age groups and dysvascular amputation more prevalent in older age groups.[5,10,12,22] Thus, it is important to examine how PPV may vary by age. Research also tends to focus on populations that are majority white, despite studies indicating that blacks have higher amputation rates.[5,12,19] This difference could be due to differences in underlying populations or differential participation.

Evaluating differences in PPV will allow us to determine if racial differences exist in the identification of LEA using EMRs.

The potential misclassification of LEA impairs our ability to accurately identify the intended population of cases and conduct research that draws meaningful conclusions from them. Furthermore, it limits our ability to perform surveillance of LEA and to accurately report the prevalence and incidence of amputation. With the implementation of the Amputation System of Care (ASoC) in 2008, the VA strengthened its commitment to "enhancing the quality and consistency of care provided to the Veteran with limb loss."[21] Without knowing how accurately we are identifying the LEA population, we are missing opportunities to understand and provide better care to this population. This misclassification may result in biased study results for those with LEA.

In this study, we aimed to estimate the positive predictive value of algorithms used to identify patients with a LEA using VHA EMRs and to determine if the positive predictive value varied by age, gender, and race. We aimed to provide recommendations for enhanced algorithm(s) for identifying LEA using VHA EMRs for a wide variety of study needs. This study is the first assessment, to our knowledge, of the accuracy of EMR-derived diagnosis and procedures codes to identify persons with LEAs.

Methods

Study Design: We conducted a validation study to determine the PPV of ICD (diagnostic and procedure) and CPT (procedure only) codes to identify LEA. Ascertainment of true LEA status was by self-report. We used data from the Corporate Data Warehouse (CDW) collected between October 1, 2005 and September 30, 2018. Subjects were eligible if they were identified as having at least one LEA diagnosis or procedure code (see Appendix for codes) during the time period listed above and were alive at the time the data were pulled. A sample of all those identified as having LEA were included in the study. We oversampled women, younger men (<40 years of age), and black men in order to make our PPV estimates in these subgroups more precise.

Data Collection: To determine “true” LEA status, we obtained self-reported LEA status via a mailed survey. Along with a letter inviting them to participate in the survey, we included an “Information sheet” (with items similar to an Informed Consent Form) that described the purpose of the study, risks and benefits, and how to opt out of further contact. The survey asked the individual to indicate yes/no to the question “Do you have a lower-extremity amputation?” for both the left and right legs. The survey also had options for amputation level (below the knee, above the knee, other). A stamped, addressed return envelope was included in the mailing to facilitate participation. Two weeks after the first mailing, we called non-responders. At the time of the call, we gave individuals the opportunity to complete the survey over the phone. We sent a second mailing to non-responders after the first call. Our final attempt at contact was a call to remaining non-responders two weeks after the second mailing. Covariate data, including age, gender, race, CPT and ICD codes, was extracted from the CDW.

Data Analysis: We conducted statistical analyses using Stata (StataCorp LP; College Station, Texas). Demographic variables (age, gender, race) and amputation-related information (number of diagnosis codes, and number of procedure codes) were summarized, stratifying by response status (Supplementary Table 1) and gender (Table 1). We calculated PPV (and 95% confidence intervals) as the quotient of the number of people identified as having LEA by the specific algorithm divided by of the number of people who self-reported LEA. We repeated this calculation, for each algorithm for each subgroup (i.e., gender, age [<40, ≥40 years] and race [black, not black]) (Tables 3-5). We estimated PPV for the current algorithm (at least one diagnosis or procedure code for LEA) as well as various other algorithms (Table 2). These algorithms were generated by examining the LEA codes from a prior study in which women reported not having an amputation yet were selected because of the presence of diagnosis and/or procedure codes. From this unpublished study we suspected that history codes (e.g., V49.75, V49.76, Z89.519, or Z89.619) might be erroneous. Thus, we created an algorithm (Algorithm 9) that did not consider a single instance of these codes as being evidence of an LEA. Thus, we sought to evaluate whether a single instance of one of these codes was likely a coding error. The other algorithms were generated based on the logic that amputees likely have more than one diagnosis code for their amputation in the EMR and that these would be recorded on different dates. We calculated false negative percentages for each algorithm, defined as the proportion of persons with an amputation not detected by the algorithm being assessed, as compared to algorithm 1.

Results

We identified a total of 32,998 individuals as eligible using the criteria of being alive and having at least one procedure or diagnosis code for LEA. 98% of eligible individuals were male. Of these, we contacted 700 individuals (Figure 1). We removed fifteen individuals from the study who were later determined to not be eligible. Fourteen of these individuals had a household member reporting them as deceased. The other individual was removed due to no procedure or diagnosis code for lower limb amputation being present in the EMR upon re-review. Of the 685 remaining individuals, we received 441 responses for a response proportion of 64% (Supplementary Table 1). Females were more likely to respond than males, as were those with a race other than black and older individuals. Those who did and did not respond had a similar number of amputation codes.

Among responders (Table 1), there was relatively even distribution of characteristics between genders. Because we oversampled men younger than 40 and black males, compared to women, the men in our sample were younger and a greater proportion were black. Females tended to have fewer codes both overall and when considering only diagnosis codes; the number of procedure codes was similar between females and males. Over half of males had over 20 codes, compared to 33% of females.

PPVs for the 9 algorithms ranged from 72% to 98% (Table 2). Algorithm 1 (which required a single diagnosis or procedure code) had the lowest PPV (72%, 95% CI: 67-76). Algorithm 8 had the highest PPV (98%, 95% CI: 95-99), but also had the highest proportion of false negatives (14%). Generally, more restrictive algorithms had both higher PPVs and higher false negative percentages.

We also assessed the performance of the 9 different algorithms among gender, race, and age subgroups (Tables 3-5). Among males, all algorithms gave high PPV estimates, ranging from 92-99%. Females had a wider distribution of estimates, ranging from 55-95%. For females, algorithm 8, requiring 3 or more procedure or three or more diagnosis codes at least one year apart, had the highest PPV estimate (95%, 95% CI: 90-99), but this algorithm also had 19% false negatives. PPV estimates ranged from 74-97% among blacks and 70-98% among those that were a race other than black. Algorithms requiring at least one procedure code or two or more diagnosis codes had high PPVs for both racial subgroups. Among those less than 40 years of age, the lowest estimate was 87% (95% CI: 79-93) for algorithm 1. Algorithms requiring at least one procedure code or two or more diagnosis codes all had PPV estimates of 99%. For those over 40 years of age, allowing any procedure or diagnosis code again gave the lowest PPV and algorithms requiring at least one procedure code or two or more diagnosis codes all gave high PPVs.

Algorithm 1, allowing any procedure or diagnosis code for LEA, consistently had the lowest PPV and algorithm 9, eliminating the four history codes but allowing any other procedure or diagnosis code for LEA, consistently had the second lowest PPV estimates across both the entire sample and all subgroups. Algorithms requiring at least one procedure code or two or more diagnosis codes generally had high PPVs, with the best performing algorithm varying among subgroups and for the entire sample. In general, increasing restriction/requirements in an algorithm resulted in a higher proportion of false negatives.

Discussion

Algorithms requiring at least one procedure code or two or more diagnosis codes gave high PPVs for the entire sample. Algorithm 8, requiring three or more procedure or three or more diagnosis codes at least one year apart, gave the highest PPV but also had the highest percentage of false negatives. In general, both the PPV estimate and the false negative percentage increased as algorithms were more restrictive. While any of the algorithms requiring at least one procedure code or two or more diagnosis codes is likely suitable, the one that is most suitable will vary depending on specific study requirements. The “best” approach for identifying the population will depend on the implications of false positives, missing a true positive, required sample size, and the quality of the specific data source used. If a large population is available or missing some individuals with an amputation is not as important, more restrictive algorithms may be suitable. If including those that do not truly have an amputation in the study does not have a large negative impact, a less specific algorithm may be suitable as it will result in a larger sample size.

Requiring a single LEA diagnosis or procedure code to generate a population of men is likely acceptable, however among women only 55% of those who had a single LEA diagnosis or procedure code endorsed having an LEA. Although requiring 3 or more procedure codes or three or more diagnosis codes at least 1 year apart gave the highest PPV for women, it is also likely not a good choice. Since the number of women with LEA who are VA patients is small (in our sample it was only 2% of the population), missing 19% of those that truly have an amputation will negatively impact the sample size. Algorithms 3-5, requiring at least one procedure code and two or more diagnosis codes a certain number of days or months apart, are the best choices for identifying a population of women, again recognizing that the best algorithm will depend on specific study needs.

Our findings suggest procedure codes are likely the most accurate way to identify LEA, but they were frequently not available. In our sample, 77% of respondents did not have a procedure code, likely because the procedure was not done at the VA and therefore the procedure code was not captured in the VHA EMR. There were no false positives for those with procedure codes. Because procedure codes are not always available it is important to take into account the number of codes, specifically diagnosis codes, and the occurrence of multiple codes on different days. Based on our results, increasing the number of codes required provided a sufficient increase in PPV. For some subpopulations, such as women, taking into account the number of codes and the occurrence of multiple codes on different days is likely necessary for an appropriate PPV. While restricting our population to individuals with multiple occurrences of the specified history codes increased the PPV estimate, the impact was not as large as anticipated likely due only a small proportion of miscodes relying only on these history codes.

One of the most important findings from our study was that PPV estimates were much lower among females. While understanding gender differences requires further investigation, we hypothesized that EMRs contain gender bias in terms of their accuracy. It is well established that EMRs contain errors and it is possible that the records of females are less accurate in general, perhaps due to the way females interact with their physicians or their care seeking patterns. Another plausible explanation is that women see different types of physicians than men, therefore resulting in differential miscoding.

A limitation to our study was that we did not have data on those that did not have an LEA diagnosis or procedure code but in fact did have an amputation, preventing us from being able to calculate sensitivity, specificity and negative predictive value (NPV). However, we were less concerned about calculating NPV (the probability that an individual who does not have a code for LEA truly does not have an LEA). Given that amputation is rare, the NPV will be high. The ability to calculate sensitivity and specificity would require collecting self-report data from a sample of all Veterans, in order to have information on true positives, false positives, true negatives, and false negatives. Ultimately, the primary aim of this study was to generate a method of accurately identifying individuals with LEA for research studies using the EMR. Given the relative rarity of LEA, an important goal is to limit the bias induced by misclassifying individuals as having an amputation when they do not. PPV is therefore the most valuable diagnostic characteristic to help achieve this aim. A second limitation was the potential for non-response bias. To reduce non-response we contacted each individual up to four times and ultimately obtained information on 64% of those approached. There were no important differences between responders and non-responders in the characteristics we measured. The magnitude of differences between groups was at most 15%, with most differences ranging from 0-6%. Due to the level of non-response bias in our sample we believe that our study population is representative of the target population.

In summary, we presented nine different algorithms with various PPV estimates and false negative percentages. We assessed PPV estimates for these algorithms performed for various subgroups and discussed the different settings in which a particular algorithm may be appropriate. These algorithms generally performed the best in men and those under the age of 40 and the differential misclassification that would result by age and sex warrants investigation in future studies. Overall and among subgroups, algorithm 1, allowing any procedure or diagnosis code for LEA, consistently had the lowest PPV and algorithm 9, eliminating the four history codes but allowing any other procedure or diagnosis code for LEA, give low PPV estimates. Algorithms requiring at least one procedure code or two or more diagnosis codes give high PPVs, and the best choice depends on specific study needs. We hope that our results will be generalizable to populations outside the VA, but futures studies should confirm this.

References

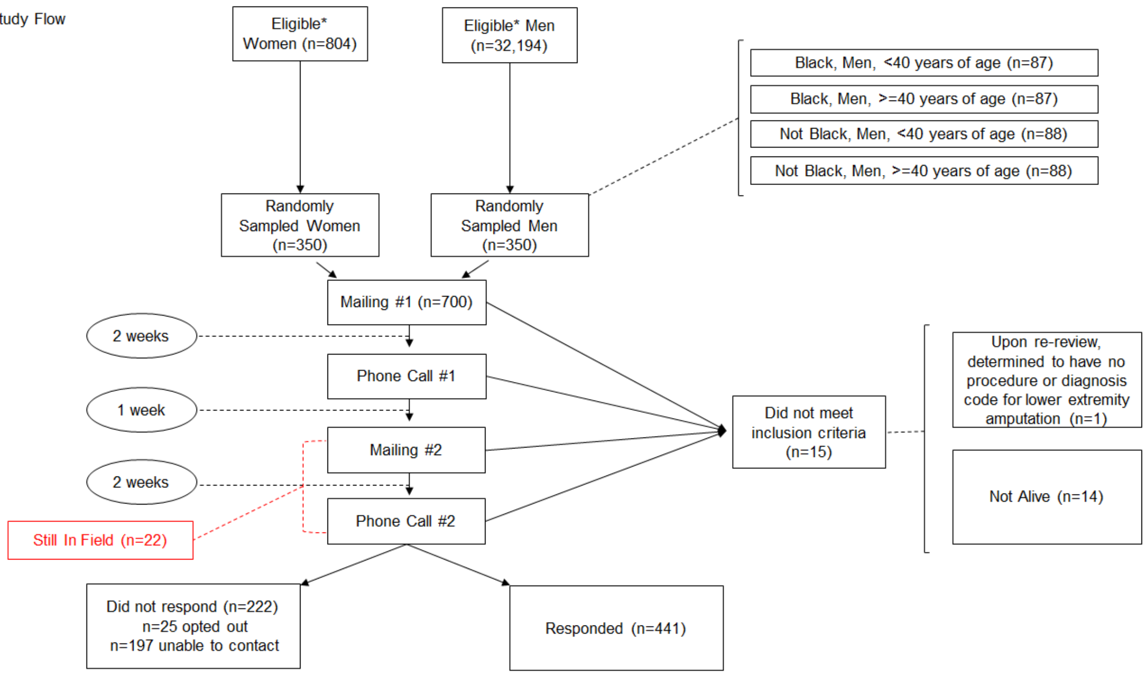
1. Ando T, Ooba N, Mochizuki M, et al. Positive predictive value of ICD-10 codes for acute myocardial infarction in Japan: a validation study at a single center. BMC Health Services Research. 2018;18:895
2. Biggerstaff K, Frankfort B, Orengo-Nania S, et al. Validity of code based algorithms to identify primary open angle glaucoma (POAG) in Veterans Affairs (VA) administrative databases. Ophthalmic Epidemiology. 2018;25(2):162-168
3. Chui S, Davis J, Giaconi J, et al. Variable validity of computer extracted problem lists for complications of diabetes mellitus within the VA Greater Los Angeles Health System. Diabetes & Metabolic Syndrome: Clinical Research & Reviews. 2017;11(2):S611-S615
4. Cochran W. *Sampling Techniques*. New York: John Wiley and Sons; 1963
5. Czerniecki J, Thompson M, Littman A, et al. Predicting reamputation risk in patients undergoing lower extremity amputation due to the complications of peripheral artery disease and/or diabetes. British Journal of Surgery. 2019; 106:1026-1034
6. Feder S, Redeker N, Sangchoon J, Schulman-Green D, et al. Validation of the ICD-9 diagnostic code for palliative care in patients hospitalized with heart failure within the Veterans Health Administration. American Journal of Hospice and Palliative Medicine. 2018; 35(7):959-965
7. Hall R, Mondor L, Porter J, Fang J, Kapral M. Accuracy of administrative data for coding of acute stroke and TIAs. Canadian Journal of Neurological Science. 2016; 43(6):765-773
8. Haneuse S, Daniels M. A general framework for considering selection bias in EHR-based studies: What data are observed and why?.Generating Evidence & Methods to improve patient outcomes. 2016; 4(1):1203

9. Holowka DW, Marx BP, Gates MA, et al. PTSD diagnostic validity in Veterans Affairs electronic records of Iraq and Afghanistan veterans. Journal of Consulting and Clinical Psychology. 2014;82(4):569-579
10. Katon J, Reiber G. Major traumatic limb loss among women veterans and servicemembers. Journal of Rehabilitation Research & Development. 2013; 50:173-182
11. Nissen F, Morales DR, Mullerova H, et al. Validation of asthma recording in the Clinical Practice Research Datalink (CPRD). BMJ Open. 2017;7(8):e017474
12. Norvell DC, Thompson ML, Boyko EJ, et al. Mortality prediction following non-traumatic amputation of the lower extremity. British Journal of Surgery. 2019; 106(7):879-888
13. Marceaux J, Soble J, O'Rourke J, et al. Validity of early-onset dementia diagnoses in VA electronic medical record administrative data. The Clinical Neuropsychologist. 2019;1-15
14. Melcer T , Sechriest VF , Walker J , Galarneau M . A comparison of health outcomes for combat amputee and limb salvage patients injured in Iraq and Afghanistan wars. Journal of Trauma and Acute Care Surgery. 2013; 75(3):S247-254
15. Michalik D, Taylor B, Panepinto J. Identification and Validation of a Sickle Cell Disease Cohort within Electronic Health Records. Academic Pediatrics. 2017;17(3):283-287
16. Miller D, Safford M, Pogach L. Who has diabetes? Best estimates of diabetes prevalence in the Department of Veterans Affairs based on computerized patient data. Diabetes Care. 2004; 27(2):B10-21
17. Omino R, Mittal S, Kramer J, Chayanupatkul M, Richardson P, Kanwal F. The validity of HCC diagnosis codes in chronic Hepatitis B patients in the Veterans Health Administration. Digestive Diseases and Sciences. 2017; 62:1180-1185
18. Rowan C, Flory J, Gerhard T, et al. Agreement and validity of electronic health record prescribing data relative to pharmacy claims data: A validation study from a US electronic health record database. Pharmacoepidemiology and Drug Safety. 2017; 26:963-972

19. Tseng C, Rajan M, Miller D, et al. Trends in initial lower extremity amputation rates among Veterans Health Administration health care system users from 2000 to 2004. Diabetes Care. 2011; 34(5): 1157-1163
20. Webster J, Crunkhorn C, Sall J, et al. Clinical practice guidelines for the rehabilitation of lower limb amputation: An update from the Department of Veterans Affairs and Department of Defense. American Journal of Physical Medicine & Rehabilitation. 2019; 98(9):820-829
21. Webster JB, Poorman CE, Cifu DX. Guest editorial: Department of Veterans Affairs Amputation System of Care: 5 Years of Accomplishments and Outcomes. Journal of Rehabilitation Research and Development. 2014; 51(4):vii-xvi
22. Ziegler-Graham K, MacKenzie EJ, Ephraim PL, et al. Estimating the prevalence of limb loss in the United States: 2005 to 2050. Archives of physical medicine and rehabilitation. 2008;89(3):422–429

Tables and Figures

Figure 1. Study Flow



*Subjects were eligible if they were identified as being alive and had at least one procedure or diagnosis code for lower limb amputation.

Table 1. Characteristics of survey respondents by gender (n=441)

	Male (n=202)		Female (n=239)	
Age (years); n(%)				
25-35	55	(27)	15	(6)
36-45	33	(16)	22	(9)
46-55	7	(3)	34	(14)
56-65	23	(10)	65	(27)
66-75	59	(29)	76	(32)
76-85	18	(9)	17	(7)
86+	7	(3)	10	(4)
Black; n(%)	90	(45)	53	(22)
Number of diagnosis and procedure codes; n(%)				
1	19	(9)	93	(39)
2-3	8	(4)	33	(14)
4-19	63	(31)	44	(18)
20+	112	(55)	80	(33)
Number of diagnosis codes; n(%)				
0	1	(0.5)	1	(0.5)
1	18	(9)	94	(39)
2-3	8	(4)	21	(9)
4-19	64	(32)	45	(19)
20+	111	(55)	78	(33)
Number of procedure codes; n(%)				
0	152	(75)	187	(78)
1	8	(4)	8	(3)
2-3	19	(9)	25	(10)
4+	23	(11)	19	(8)

Table 2. Positive predictive value (PPV) of algorithms used to identify lower extremity amputation in Veterans Health Administration electronic health records with a self-reported gold standard

Number	Algorithm	Algorithm Positive*	True Amputation	PPV (%)	95% CI	False Negative (%)**
1	Any code	441	316	72	67-76	0
2	At least one procedure code or two or more diagnosis codes	330	310	94	91-96	2
3	At least one procedure code or two or more diagnosis codes on different days	325	309	95	92-97	2
4	At least one procedure code or two or more diagnosis codes at least 30 days apart	315	303	96	93-98	4
5	Two or more procedure or 2 or more diagnosis codes at least 30 days apart	314	302	96	93-98	5
6	Two or more procedure or 2 or more diagnosis codes at least 1 year apart	289	280	97	94-99	11
7	Three or more procedure or 3 or more diagnosis codes at least 30 days apart	304	296	97	95-99	6
8	Three or more procedure or three or more diagnosis codes at least 1 year apart	277	271	98	95-99	14
9	Presence of any code excluding codes V49.75, V49.76, Z89.519, Z89.619 if appears once***	348	296	85	81-89	6

*Algorithm positive is defined as meeting the inclusion criteria of being alive and meeting the criteria for the specified algorithm
**Calculated based on algorithm 1
***Indicated codes are for acquired amputation above or below the knee for both ICD-9 and ICD-10

Table 3. Positive predictive value (PPV) of algorithms used to identify lower extremity amputation in Veterans Health Administration electronic health records with a self-reported gold standard stratified by gender

Algorithm Number*	Gender	Algorithm Positive**	True Amputation	PPV (%)	95% CI	False Negative (%)***
1	Male	202	185	92	87-95	0
1	Female	239	131	55	48-61	0
2	Male	184	182	99	96-100	2
2	Female	146	128	88	81-93	2
3	Male	184	182	99	96-100	2
3	Female	141	127	90	84-94	3
4	Male	182	181	99	97-100	2
4	Female	133	122	92	86-96	7
5	Male	181	180	99	97-100	3
5	Female	133	122	92	86-96	7
6	Male	169	168	99	97-100	9
6	Female	120	112	93	87-97	15
7	Male	179	178	99	97-100	4
7	Female	125	118	94	89-98	10
8	Male	166	165	99	97-100	11
8	Female	111	106	95	90-99	19
9	Male	179	172	96	92-98	7
9	Female	169	124	73	66-80	5

*Refer to Table 3 for algorithm descriptions
**Algorithm positive is defined as meeting the inclusion criteria of being alive and meeting the criteria for the specified algorithm
***Calculated based on algorithm 1

Table 4. Positive predictive value (PPV) of algorithms used to identify lower extremity amputation in Veterans Health Administration electronic health records with a self-reported gold standard stratified by race

Algorithm Number*	Race	Algorithm Positive**	True Amputation	PPV (%)	95% CI	False Negative (%)***
1	Black	143	106	74	66-81	0
1	Race other than black	298	210	70	65-76	0
2	Black	109	104	95	90-98	2
2	Race other than black	221	206	93	89-96	2
3	Black	109	104	95	90-98	2
3	Race other than black	216	205	95	91-97	2
4	Black	107	103	96	91-99	3
4	Race other than black	208	200	96	93-98	5
5	Black	106	102	96	91-99	4
5	Race other than black	208	200	96	93-98	5
6	Black	98	94	96	90-99	11
6	Race other than black	191	186	97	94-99	11
7	Black	104	101	97	92-99	5
7	Race other than black	200	195	98	94-99	7
8	Black	94	91	97	91-99	14
8	Race other than black	183	180	98	95-100	14
9	Black	120	101	84	76-90	5
9	Race other than black	228	195	86	80-90	7

*Refer to Table 3 for algorithm descriptions
***Algorithm positive is defined as meeting the inclusion criteria of being alive and meeting the criteria for the specified algorithm
***Calculated based on algorithm 1

Table 5. Positive predictive value (PPV) of algorithms used to identify lower extremity amputation in Veterans Health Administration electronic health records with a self-reported gold standard stratified by age

Algorithm Number*	Age	Algorithm Positive**	True Amputation	PPV (%)	95% CI	False Negative (%)***
1	<40 years of age	106	92	87	79-93	0
1	≥40 years of age	335	224	67	61-72	0
2	<40 years of age	91	90	99	94-100	2
2	≥40 years of age	239	220	92	88-95	2
3	<40 years of age	91	90	99	94-100	2
3	≥40 years of age	234	219	94	90-96	2
4	<40 years of age	91	90	99	94-100	2
4	≥40 years of age	224	213	95	91-98	5
5	<40 years of age	91	90	99	94-100	2
5	≥40 years of age	223	212	95	91-98	5
6	<40 years of age	87	86	99	94-100	7
6	≥40 years of age	202	194	96	92-98	13
7	<40 years of age	89	88	99	94-100	4
7	≥40 years of age	215	208	97	93-99	7
8	<40 years of age	83	82	99	93-100	11
8	≥40 years of age	194	189	97	94-99	16
9	<40 years of age	91	83	91	83-96	10
9	≥40 years of age	257	213	83	78-87	5

*Refer to Table 3 for algorithm descriptions
***Algorithm positive is defined as meeting the inclusion criteria of being alive and meeting the criteria for the specified algorithm
***Calculated based on algorithm 1

Supplementary Table 1. Characteristics of study population by response status

	Responders (n=441)		Non-Responders (n=244)	
Male; n(%)	202	(46)	139	(57)
Age (years); n(%)				
25-35	70	(15)	64	(26)
36-45	55	(12)	44	(18)
46-55	41	(9)	27	(11)
56-65	88	(20)	45	(18)
66-75	135	(31)	40	(16)
76-85	35	(8)	16	(7)
86+	17	(4)	8	(3)
Black; n(%)	143	(32)	114	(47)
Number of diagnosis and procedure codes; n(%)				
1	112	(25)	53	(22)
2-3	30	(7)	18	(7)
4-19	107	(24)	60	(25)
20+	192	(44)	113	(46)
Number of diagnosis codes; n(%)				
0	2	(0.5)	1	(0.5)
1	112	(25)	52	(21)
2-3	29	(7)	18	(7)
4-19	109	(25)	61	(25)
20+	189	(43)	112	(46)
Number of procedure codes; n(%)				
0	339	(77)	195	(80)
1	16	(4)	10	(4)
2-3	44	(10)	16	(7)
4+	42	(10)	23	(9)