



HHS Public Access

Author manuscript

J Vasc Surg. Author manuscript; available in PMC 2023 July 01.

Published in final edited form as:

J Vasc Surg. 2022 July ; 76(1): 266–271.e2. doi:10.1016/j.jvs.2022.01.132.

Validation of an indirect linkage algorithm to combine registry data with Medicare claims R1WC: 331/2546

Jialin Mao, MD, MS¹, Kayla O. Moore, MS², Jesse A. Columbo, MD, MS³, Kunal S. Mehta, MD³, Philip P. Goodney, MD, MS^{3,*}, Art Sedrakyan, MD, PhD^{1,*}

¹Department of Population Health Sciences, Weill Cornell Medicine, New York, NY

²Heart & Vascular Center, Dartmouth-Hitchcock Medical Center, Lebanon, NH

³Section of Vascular Surgery, Dartmouth-Hitchcock Medical Center, Lebanon, NH

Abstract

Introduction—Linkage of registries to Medicare claims data can help extend follow-up for patients receiving medical devices. This study sought to test and validate an algorithm that does not require patient identifiers to link a national vascular registry and Medicare claims data.

Methods—We used data from the Vascular Quality Initiative (VQI), a registry capturing data from more than 600 centers on several different vascular procedures and Medicare claims from 2003 to 2018. We restricted to patients aged 65 years and older who had fee-for-service entitlement at the time of the procedure for this study. We performed an indirect linkage to combine VQI with Medicare at the patient level using a sequential algorithm based on patient's date of birth, sex, zip code, procedure date, and procedure facility. We compared this against a gold standard of a cohort directly linked using social security numbers (SSNs). We calculated the matching rate and accuracy overall, and before and after October 2015 when the ICD-10 system was adopted in the US.

Results—A total of 144,045 VQI-Medicare linked patients were in the gold standard cohort. Using the indirect linking algorithm, we matched 133,966 of the 144,045 VQI patients to their Medicare claims (matching rate: 93.0%). Among these, 133,104 patients were correctly matched (matching accuracy: 99.4%). The matching rate was higher when the indirect linkage was implemented in ICD-10 coded data than in ICD-9 coded data (94.0% vs. 92.2%). Accuracy of the indirect linkage remained high for all procedure modules post-ICD-10 coding change (overall 99.4%, range 99.0–99.7%).

Conclusion—In this study, we successfully used indirect identifiers to link the VQI to Medicare claims with more than 90% success and more than 99% accuracy. When direct linkage of registry-

Correspondence: Jialin Mao, MD, MS, Assistant Professor of Population Health Sciences, Weill Cornell Medicine, 570 Lexington Ave, 10th Floor, New York, NY 10022, Tel: 646-962-8030, Fax: 646-962-0281, jim2012@med.cornell.edu.

*Co-senior authors

From the Vascular Quality Initiative - Vascular Implant Surveillance and Interventional Outcome Network (VQI-VISION)

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

claims data using SSNs is not possible because of availability, confidentiality, or both, this method for indirect linkage provides a suitable alternative. The matching rate and accuracy help ensure the accuracy of long-term follow-up and the completeness and representativeness of linked databases for relevant research and quality improvement initiatives.

Manuscript summary

This study demonstrated a 93% matching rate and more than 99% accuracy of an indirect linkage algorithm to link the Vascular Quality Initiative Registry to Medicare claims. The validity of this method helps ensure the accuracy of long-term follow-up and the completeness and representativeness of linked databases for relevant research and quality improvement initiatives.

Keywords

Vascular surgery; medical device research; registry study; claims linkage

Introduction

In the past two decades, there has been a rapid increase in the number of medical devices available to patients.¹ In particular, the adoption of endovascular devices in the treatment of cardiovascular disease has increased substantially.^{2–7} Millions of patients who received these procedures have benefited from improved disease symptoms and greater longevity. Careful monitoring of patients' outcomes after cardiovascular procedures is vital to safe patient care. To achieve this, many societies have employed the use of medical and procedural registries. These registries capture granular patient-level data and allow the study of procedural results that form the foundation for safe and effective patient care.

However, registry data have limitations. Notably, the collection of long-term patient follow-up is labor-intensive and challenging due to substantial loss to follow-up. Therefore, some investigators have proposed the linkage of registries to administrative claims data to improve patient follow-up in a less labor-intensive manner. The linkage of registry and claims data can take place using identifiable patient information, such as social security numbers (SSNs), or using indirect identifiers, such as patient age, sex, and procedure date, when the use of direct identifiers is not possible either due to confidentiality, availability, or both.

The Vascular Implant Surveillance and Interventional Outcomes (VISION) Coordinated Registry Network (CRN)⁸ is a partnership between the Society for Vascular Surgery (SVS) Patient Safety Organization (PSO) and the Medical Device Epidemiology Network (MDEpiNet), supported by the US Food and Drug Administration (FDA), to improve the assessment of long-term patient outcomes following vascular procedures. Currently, VISION employs a combined direct and indirect linkage methodology to link the Vascular Quality Initiative (VQI) registry and Medicare claims. This approach retains a more complete cohort of patients for vascular research and quality improvement initiatives than direct linkage alone. However, the absolute matching rate and accuracy of the indirect linkage remain uncertain. The indirect linkage algorithm has shown a high matching rate and accuracy in a limited orthopedic registry cohort linked to New York State data,⁹ but has never been tested in a large, national-scale linkage with Medicare data.

The objective of the current study is to test and validate the indirect linkage algorithm that does not require direct patient identifiers in the linkage of a national vascular registry and Medicare claims. We compared the indirectly linked cohort against the cohort linked based on the gold standard using SSNs and determined the matching rate and accuracy of the indirect linkage. Additionally, with the transition from ICD-9 to ICD-10 codes in October 2015, we assessed the linkage performance before and after the transition. In doing so, we provide investigators with a guide on the method of the indirect linkage between registry and claims data and the validity of this method.

Methods

Human subjects protection

The VISION CRN follows strict guidelines to protect research-identifiable data. Data use agreements for the data linkage and analytical work have been established between the MDEpiNet coordinating center at Weill Cornell Medicine, the Centers for Medicare & Medicaid Services (CMS), and the Society for Vascular Surgery Patient Safety Organization (SVS PSO). All protected health information (PHI) data are stored on a HIPAA and FISMA compliant institutional secured server, accessible only to study personnel with unique login credentials. The study protocol has been approved by the Weill Cornell Medicine Institutional Review Board.

Data sources

We used the VQI registry and Medicare fee-for-service (FFS) claims data for this validation study. The VQI is a national vascular registry that collects clinical, demographic, procedural, and one-year outcomes data on patients undergoing vascular procedures from over 600 academic and community hospitals.¹⁰ We extracted patients' Medicare claims from the Medicare Provider and Analysis Review (MedPAR), outpatient, and Part B claims data. Patients' FFS entitlement status was determined from the Master Beneficiary Summary File.

Gold standard cohort creation

Our gold standard cohort included VQI patients treated at US hospitals between 2003 and 2018, who were successfully linked to Medicare using direct identifiers (name and social security number) by the CMS data vendor. Seven VQI registry modules were included: endovascular aortic aneurysm repair (EVAR), open aortic aneurysm repair (OAR), thoracic endovascular aortic aneurysm repair (TEVAR), carotid endarterectomy (CEA), carotid artery stenting (CAS), lower extremity (suprainguinal and infrainguinal) bypass (LEB), and peripheral vascular intervention (PVI). We restricted the cohort to patients aged above 65 years for the study. Because CMS data does not include all claims for Medicare Advantage patients (detailed explanation in Appendix), the validation of the algorithm was performed among patients who had FFS entitlement at the time of the procedure. FFS entitlement was defined as having both Part A and B enrollment and no Medicare Advantage enrollment. The cohort creation and validation process are depicted in Figure 1.

The direct linkage was performed by the CMS using direct patient identifiers (i.e., Social security number (SSN)), which has been previously reported.¹¹ The MDEpiNet coordinating

center received a crosswalk of synthetic Medicare beneficiary and registry identifiers (i.e., a unique number assigned to the patient that does not reveal the patient's real identifiers, e.g., name, SSN, or medical record number, and cannot be tracked back to the individual) from the CMS, leaving out the direct identifiers. Patients with inconsistent matches (e.g., one patient to more than one claims file) were removed.

Indirect linkage procedure

Eligible records for indirect linkage were prepared from the registry and Medicare data. The coordinating center received the registry data from the VQI with the synthetic patient and procedure identifiers and used these identifiers to extract relevant registry records. Medicare records were queried to identify the vascular procedures listed above using a set of procedure trigger codes. All procedures were identified from MedPAR data using ICD procedure codes. In addition, outpatient PVI procedures were identified from outpatient and Part B claims using CPT procedure codes.

Using the eligible records from registry and Medicare claims data, the indirect linkage was conducted using a sequential algorithm (Table 1).^{9, 12} The indirect identifiers used were patient's date of birth, sex, zip code (first 3 digits), procedure date, and procedure facility. The first step was to match with all five variables. The algorithm then accounted for missing and inaccurate variables by allowing some flexibility in dates of birth and procedure and indirect identifiers used. During the developmental stage, we tested a few iterations of algorithms to decide the flexibility given to these variables (Appendix). Flexibility was allowed for no more than two matching variables simultaneously to achieve additional matching with minimally reduced accuracy. After completing all matching steps, all inconsistent matches were excluded (e.g., one registry patient linked to two Medicare beneficiaries) and considered non-match

Validation against the gold standard

Validation of the indirect linkage was performed by examining the matching rate and accuracy of matching. The matching rate was defined as the proportion of records linked by the indirect algorithm among all records matched directly with SSNs. Accuracy was defined as the proportion of correct matches out of all indirectly linked records, using the direct linkage as a gold standard.

$$\begin{aligned} \text{Matching rate} &= \frac{\text{Number of VQI records matched to Medicare with indirect identifiers}}{\text{Number of VQI records matched to Medicare with direct identifiers}} \\ \text{Accuracy} &= \frac{\text{Number of indirectly matched records that were correct matches}}{\text{Number of VQI records matched to Medicare with indirect identifiers}} \end{aligned}$$

Due to the nature of linkage, we could not use regular bootstrapping that performs repeat sampling with replacement. Thus, to estimate 95% confidence intervals of the matching rate and accuracy, we performed repeat random sampling (N=500) of eligible registry records with a 5% dropout rate. For every iteration, a random 5% sample of registry data was dropped, and linkage was performed for the remaining 95%. We obtained the 2.5th and

97.5th percentiles of the matching rate and accuracy estimated from this repeat sampling. We also examined the matching rate and accuracy of the algorithm before and after the ICD-10 transition (October 2015). We also examined the unmatched records to explore reasons for non-match. All analyses were performed using SAS 9.4 (Cary, NC).

Results

After excluding inconsistent matches, the directly linked VQI-Medicare cohort included 217,640 patients aged 65 years or older at the time of the procedure. The mean age at the time of procedure was 75.0 (SD 6.9) years and 36.9% were female. We removed 73,595 patients that did not have FFS Medicare entitlement. This left 144,045 patients who had Medicare FFS entitlement at the time of the procedure in the gold standard cohort.

Using the indirect linkage algorithm, we matched 133,966 of the 144,045 patients (matching rate: 93.0%). Among these, 133,104 patients were correctly matched (accuracy: 99.4%). The matching rate of the indirect linkage varied across procedure modules (Table 2). The rate was highest for EVAR (93.0%), CEA (95.0%), and PVI (93.6%), and lowest for CAS (87.1%). The accuracy of the indirect linkage was high for all procedures, ranging between 98.9% and 99.7%.

Over 40% of the validation cohort was accrued in the ICD-10 era (ICD-9: N=79,411 procedures, ICD-10: N=64,634 procedures). The matching rate was higher when the indirect linkage was implemented in ICD-10 coded data than in ICD-9 coded data (94.0% vs. 92.2%), most notably for CAS (89.5% vs. 83.7%), LEB (94.2% vs. 88.3%), and TEVAR (95.4% vs. 88.7%) (Figure 2). Accuracy of the indirect linkage remained high for all procedure modules following the ICD-10 coding transition (overall 99.4%, range 99.0–99.7%).

In total, 10,079 directly linked patients were not matched in the indirect linkage. Of them, 4,500 (44.6%) did not have Medicare claims within nine days before or after their VQI procedure dates and 3,738 (37.1%) had claims within the specified date range, but no procedure trigger codes were identified. Fewer unmatched cases were related to the absence of claims in the ICD-10 era than in the ICD-9 era (42.4% vs. 46.0%). The remaining 1,841 (18.3%) patients were not matched due to differences in other indirect identifiers or were matched to more than one Medicare beneficiary.

Discussion

In this study, we tested and validated an indirect linkage algorithm to combine registry data from the VQI with claims data from Medicare at the patient level, without the use of direct identifiers, such as name and SSN. Compared with our gold standard of a directly linked cohort, the indirect linkage algorithm showed an overall 93% matching rate and greater than 99% accuracy. Furthermore, these findings were even more robust in ICD-10 data, where the matching rate improved for all procedure modules and the accuracy remained high.

This study demonstrated the validity of the indirect linkage algorithm in a large patient cohort. The accuracy of the linkage algorithm remained similar across VQI modules and

was similar to that shown in the previous small-scale orthopedic cohort.⁹ This means that by using the indirect linkage, researchers can find the correct follow-up for registry patients >99% of the time. In addition, this study showed that the algorithm remained reliable after the transition from the ICD-9 to the ICD-10 coding system. These results demonstrated that the indirect linkage algorithm is generalizable and applicable to different disease areas and databases, before and after the ICD coding system transition.

The matching rate of the algorithm was over 90% both in the Medicare setting (93% overall in this study) as well as in the setting of all-payer state data (92% in the previous study⁹). The high matching rate helps ensure that the linked cohort accurately represents the cohort in the initial unlinked dataset. In addition, the current study showed that the matching rate was procedure-dependent, as it varied across VQI modules, ranging from 87% to 95%. This variation may be partially related to the sensitivity of trigger codes and the completeness of claims submission. The matching rate of the indirect linkage was slightly higher in the ICD-10 era than in the ICD-9 era. Notably, a lower proportion of patients in the ICD-10 era had no claims found within the specified time window. The claims submission seemed more complete in the ICD-10 era, and the matching rate, as a result, improved.

The VISION CRN currently adopts a hybrid approach for the VQI-Medicare linkage, combining direct linkage with an indirect linkage for centers that only share indirect identifiers and for records for which direct identifiers were not available. In the most recent years (2017–2018) of the current linkage, approximately 40% of the VQI records did not have direct identifiers available or shared. The combined use of direct and indirect linkages helps maximize the number of patients being matched to Medicare claims, thus increasing the representativeness of the linked data. The accuracy of the indirect algorithm ensures the reliability of the linked data. At the time of writing this manuscript, 192,469 VQI patients who had FFS status were matched to Medicare in the latest linkage (data up to December 31, 2018) using this hybrid approach. This number continues to evolve as new data become available. The most updated information is available on the SVS VQI website (www.vqi.org).

The implementation of linkage between the VQI registry and claims data has important implications in vascular quality improvement, outcomes research, and medical device surveillance. It combines the granular clinical details collected in the registry and longitudinal data from claims data. The VQI-Medicare linkage enables long-term follow-up of patients for key endpoints, such as mortality, reintervention, and imaging surveillance. For example, previous research has shown that claims data can be used to reliably assess reinterventions following initial aortic aneurysm repairs.^{13, 14} In addition, Medicare claims data can also help understand the costs of index vascular procedures and reinterventions during follow-up.^{15, 16} The addition of these long-term endpoints can help enrich quality improvement feedback that the registry provides to participating centers on center-specific and device-specific patient outcomes. The linked data can also help advance vascular outcomes research and device surveillance by providing important evidence of the safety and effectiveness of contemporary treatments beyond the first year after initial procedures. Multiple stakeholders stand to benefit from this efficient, comprehensive, and cost-effective resource. Evidence derived from this data source can help regulators and

manufacturers better understand vascular device performance in the real world, facilitate clinical decision-making, and improve patient safety.

There are a few limitations to the current study. The SSN-BeneID crosswalk generated by CMS direct linkage was used as the gold standard. However, there might be some inaccuracies in the direct linkage. But the inconsistent match seen in the data pre-processing was very low (1%), and therefore, would not change our estimates significantly. We did not calculate the linkage and cohort creation estimates for Medicare beneficiaries aged under 65 years, mostly end-stage renal disease or disability patients, because the denominator of Medicare-eligible patients was difficult to determine. We do not expect the matching rate and accuracy of the algorithm to be significantly different in a cohort of patients under 65 years in Medicare. The final number of patients that can be linked for research in this age group would be dependent on the number of Medicare-eligible patients.

In conclusion, in this study, we successfully used indirect identifiers to link the VQI to Medicare claims with more than 90% success and more than 99% accuracy. When direct linkage of registry-claims data using social security numbers is not possible because of availability, confidentiality, or both, this method for indirect linkage provides a suitable alternative. The matching rate and accuracy help ensure the accuracy of long-term follow-up, completeness, and representativeness of linked databases for relevant research and quality improvement initiatives.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding/Support

This study was supported by the Office of the Assistant Secretary for Planning and Evaluation Patient-Centered Outcomes Research Trust Fund under Interagency Agreement (#750119PE060048), through the US Food and Drug Administration (Grant number U01FD006936, PI: AS).

Disclosures

JM receives funding from NHLBI (K01HL159315), which supports her effort.

References

1. Darrow JJ, Avorn J, Kesselheim AS. FDA Regulation and Approval of Medical Devices: 1976–2020. *JAMA*. 2021;326(5):420–32. [PubMed: 34342614]
2. Suckow BD, Goodney PP, Columbo JA, Kang R, Stone DH, Sedrakyan A, et al. National trends in open surgical, endovascular, and branched-fenestrated endovascular aortic aneurysm repair in Medicare patients. *J Vasc Surg*. 2018;67(6):1690–7 e1. [PubMed: 29290495]
3. Wang GJ, Jackson BM, Foley PJ, Damrauer SM, Goodney PP, Kelz RR, et al. National trends in admissions, repair, and mortality for thoracic aortic aneurysm and type B dissection in the National Inpatient Sample. *J Vasc Surg*. 2018;67(6):1649–58. [PubMed: 29506945]
4. Lichtman JH, Jones MR, Leifheit EC, Sheffet AJ, Howard G, Lal BK, et al. Carotid Endarterectomy and Carotid Artery Stenting in the US Medicare Population, 1999–2014. *JAMA*. 2017;318(11):1035–46. [PubMed: 28975306]

5. Goodney PP, Beck AW, Nagle J, Welch HG, Zwolak RM. National trends in lower extremity bypass surgery, endovascular interventions, and major amputations. *J Vasc Surg.* 2009;50(1):54–60. [PubMed: 19481407]
6. Masoudi FA, Ponirakis A, de Lemos JA, Jollis JG, Kremers M, Messenger JC, et al. Trends in U.S. Cardiovascular Care: 2016 Report From 4 ACC National Cardiovascular Data Registries. *J Am Coll Cardiol.* 2017;69(11):1427–50. [PubMed: 28025065]
7. Culler SD, Cohen DJ, Brown PP, Kugelmass AD, Reynolds MR, Ambrose K, et al. Trends in Aortic Valve Replacement Procedures Between 2009 and 2015: Has Transcatheter Aortic Valve Replacement Made a Difference? *Ann Thorac Surg.* 2018;105(4):1137–43. [PubMed: 29394992]
8. Tsougranis G, Eldrup-Jorgensen J, Bertges D, Schermerhorn M, Morales P, Williams S, et al. The Vascular Implant Surveillance and Interventional Outcomes (VISION) Coordinated Registry Network: An effort to advance evidence evaluation for vascular devices. *J Vasc Surg.* 2020;72(6):2153–60. [PubMed: 32442604]
9. Mao J, Etkin CD, Lewallen DG, Sedrakyan A. Creation and Validation of Linkage Between Orthopedic Registry and Administrative Data Using Indirect Identifiers. *J Arthroplasty.* 2019;34(6):1076–81 e0. [PubMed: 30803801]
10. Cronenwett JL, Kraiss LW, Cambria RP. The Society for Vascular Surgery Vascular Quality Initiative. *J Vasc Surg.* 2012;55(5):1529–37. [PubMed: 22542349]
11. Holmes DR Jr., Brennan JM, Rumsfeld JS, Dai D, O'Brien SM, Vemulapalli S, et al. Clinical outcomes at 1 year following transcatheter aortic valve replacement. *JAMA.* 2015;313(10):1019–28. [PubMed: 25756438]
12. Hoel AW, Faerber AE, Moore KO, Ramkumar N, Brooke BS, Scali ST, et al. A pilot study for long-term outcome assessment after aortic aneurysm repair using Vascular Quality Initiative data matched to Medicare claims. *J Vasc Surg.* 2017;66(3):751–9 e1. [PubMed: 28222989]
13. Columbo JA, Kang R, Hoel AW, Kang J, Leinweber KA, Tauber KS, et al. A comparison of reintervention rates after endovascular aneurysm repair between the Vascular Quality Initiative registry, Medicare claims, and chart review. *J Vasc Surg.* 2019;69(1):74–9 e6. [PubMed: 29914838]
14. Columbo JA, Sedrakyan A, Mao J, Hoel AW, Trooboff SW, Kang R, et al. Claims-based surveillance for reintervention after endovascular aneurysm repair among non-Medicare patients. *J Vasc Surg.* 2019;70(3):741–7. [PubMed: 30922744]
15. Columbo JA, Goodney PP, Gladders BH, Tsougranis G, Wanken ZJ, Trooboff SW, et al. Medicare costs for endovascular abdominal aortic aneurysm treatment in the Vascular Quality Initiative. *J Vasc Surg.* 2021;73(3):1056–61. [PubMed: 32682064]
16. Trooboff SW, Wanken ZJ, Gladders B, Lucas BP, Moore KO, Barnes JA, et al. Characterizing Reimbursements for Medicare Patients Receiving Endovascular Abdominal Aortic Aneurysm Repair at Vascular Quality Initiative Centers. *Ann Vasc Surg.* 2020;62:148–58. [PubMed: 31610277]

Article Highlight

Type of Research:

Retrospective study of prospectively collected registry data

Key Findings:

In this study, we successfully used indirect identifiers to link the Vascular Quality Initiative Registry to Medicare claims with a 93% matching rate and more than 99% accuracy, compared to a gold standard of direct linkage.

Take home Message:

The high matching rate and accuracy help ensure the accuracy of long-term follow-up and the completeness and representativeness of linked databases for relevant research and quality improvement initiatives.

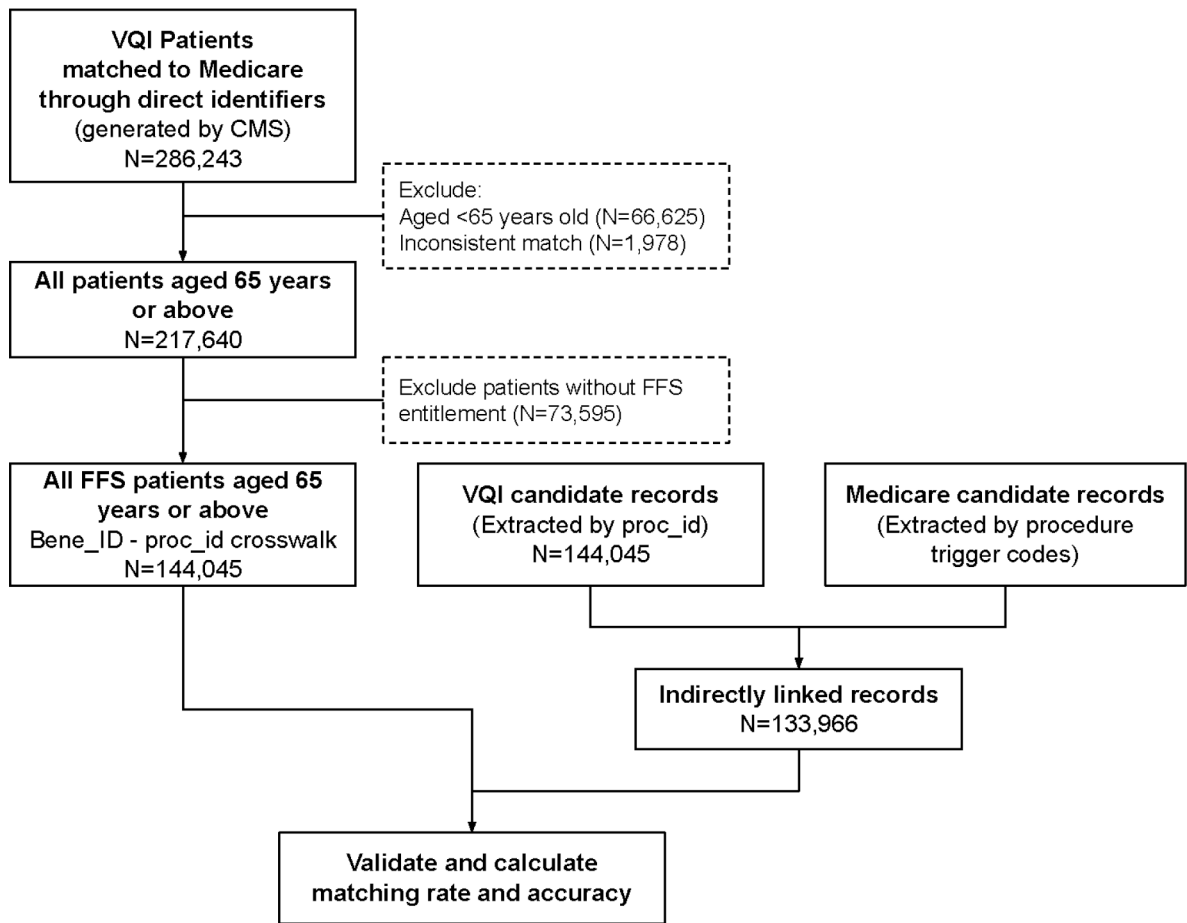


Figure 1. Overview of the cohort creation and validation process.

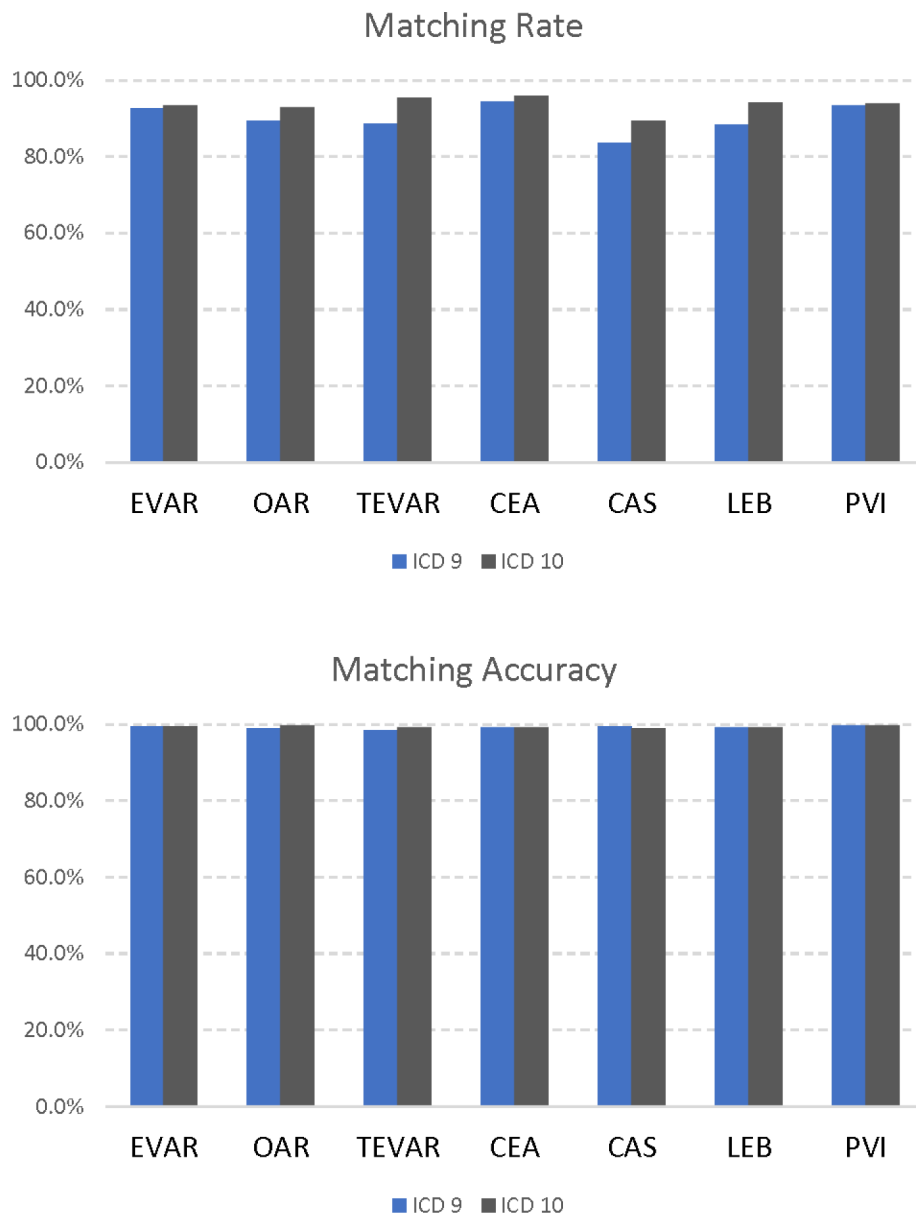


Figure 2. Linkage matching rate and accuracy for ICD-9 (black) and ICD-10 (grey) coded data.

Table 1.

Sequential algorithm for the indirect linkage between VQI registry and Medicare claims data.

Step	Center ID	Date of birth	Sex	Procedure date	Zip code
1	x	x	x	x	x
2	x	x	x	x	
3	x	x	x	(+/-) 9	
4	x	x		x	x
5	x	x		x	
6	x	(+/-) 3	x	x	x
7	State	x	x	x	x

The first step matched with all five variables. The algorithm then accounted for missing and inaccurate variables by allowing some flexibility in dates of birth and procedure and indirect identifiers used.

Table 2.

Linkage matching rate and accuracy for each procedure module among Fee-for-service Medicare beneficiaries aged ≥ 65 years old.

Procedure Module	Time range	N Directly linked	N indirectly linked	N accurately linked	Matching rate (95% CI)	Accuracy (95% CI)
EVAR	2003–18	17810	16562	16451	93.0% (92.9–93.1%)	99.3% (99.3–99.4%)
OAR	2003–18	4311	3895	3863	90.4% (90.2–90.5%)	99.2% (99.1–99.2%)
TEVAR	2010–18	3489	3225	3191	92.4% (92.2–92.6%)	98.9% (98.9–99.0%)
CEA	2003–18	40465	38441	38107	95.0% (95.0–95.0%)	99.1% (99.1–99.2%)
CAS	2005–18	8947	7796	7733	87.1% (87.0–87.3%)	99.2% (99.2–99.2%)
LEB	2003–18	17771	16074	15951	90.5% (90.4–90.6%)	99.2% (99.2–99.3%)
PVI	2004–18	51252	47973	47808	93.6% (93.6–93.7%)	99.7% (99.6–99.7%)