BACKGROUND: On October 1, 2015, the Center for Medicare and Medicaid Services transitioned from the International Classification of Diseases, Ninth Revision (ICD-9) to the ICD, Tenth Revision (ICD-10) compendium of codes for diagnosis and billing in health care, but translation between the two is often inexact. Here we describe a validated crosswalk to translate ICD-9 codes into ICD-10 codes, with a focus on complications after carotid revascularization and endovascular aortic aneurysm repair.

METHODS AND RESULTS: We devised an 8-step process to derive and validate ICD-10 codes from existing ICD-9 codes. We used publicly available sources, including the General Equivalence Mapping database, to translate ICD-9 codes used in prior work to ICD-10 codes. We defined ICD-10 codes as validated if they were concordant with the initial ICD-9 codes after manual comparison by two physicians. Our primary validation measure was the percent of valid ICD-10 codes out of the total ICD-10 codes obtained during translation. We began with 126 ICD-9 diagnosis codes used for complication identification after carotid revascularization procedures, and 97 ICD-9 codes for complications after endovascular aortic aneurysm procedures. Translation generated 143 ICD-10 codes for carotid revascularization, a 14% increase from the initial 126 codes. Manual comparison demonstrated 98% concordance, with 99% agreement between the reviewers. Similarly, we identified 108 ICD-10 codes for endovascular aortic aneurysm repair, an 11% increase from the initial 97 ICD-9 codes. We again noted excellent concordance and agreement (98% and 100%, respectively). Manual review identified 4 ICD-10 codes incorrectly translated from ICD-9 codes for carotid revascularization, and 3 codes incorrectly translated for endovascular aortic aneurysm repair.

CONCLUSIONS: Algorithms to crosswalk lists of ICD-9 codes to ICD-10 codes can leverage electronic resources to minimize the burden of code translation. However, manual review for code validation may be necessary, with collaboration across institutions for researchers to share their efforts.
Insurance claims-based research is an integral component of cardiovascular knowledge generation. The most commonly used insurance claims coding system, the International Classification of Diseases, Ninth Revision (ICD-9), has been in use for over 3 decades, and during this time period the use of claims-based research has increased dramatically. Investigators have used claims data to contribute to many important health care and policy advances in cardiovascular disease and many other medical fields.

The application of claims data to research is predicated on the use of billing codes as a proxy for the study of clinical events. The generation and validation of algorithms to define the codes, which represent actual clinical events is an essential step in high-quality outcomes research. Many stakeholders participate in this important step, including the Agency for Healthcare Research and Quality, the Center for Medicare and Medicaid Services, and independent investigators who publish their research findings, as well as the coding algorithms which form the foundation of their work. By leveraging the efforts from these sources, robust repositories of ICD-9 codes that can be reliably used in clinical research have been created.

In October 2015, the Center for Medicare and Medicaid Services transitioned from ICD-9 to the ICD, Tenth Revision (ICD-10) list of billing codes. This transition expanded the number of available diagnosis codes from ≈14000 in ICD-9 to over 69000 with ICD-10. This change created several challenges. First, validated lists of ICD-9 codes previously used to identify specific outcomes can no longer be used with ICD-10 data. Second, there is infrequently an exact match between an individual ICD-9 code and an individual ICD-10 code. Computer-based algorithms have been designed to translate ICD-9 into ICD-10 codes, such as the General Equivalence Mapping (GEM) database, but these resources have not been tested for the purpose of clinical research.

Therefore, it was our objective to use publicly available sources to validate a crosswalk by which a list of known ICD-9 codes could be translated into a list of ICD-10 codes. We studied codes used extensively in our research program as part of a longitudinal effort in cardiovascular outcomes research, with the belief that any lessons learned would apply more broadly to ICD-9 to ICD-10 transition efforts.

METHODS

Analytic Overview of Our 8-Step Validation Process

This report describes the analytic methods used by our research group for ICD-9 to ICD-10 code translation. The data and materials used for this report are cited and publicly available in the hopes that it may serve as a resource, and that others may reproduce and replicate the procedure.

We devised an 8-step process to derive and validate ICD-10 codes from an existing set of ICD-9 codes representing outcomes across several body systems (Figure 1). This process was developed in an iterative fashion with input from all authors and shared with collaborators as part of an ongoing National Institute on Aging Program Project (P01-AG019783).

The first step began with a group of ICD-9 codes and their corresponding labels that had been validated for outcome analysis. In the second step, we translated the ICD-9 codes to ICD-10 codes using the GEM database, and then linked the ICD-10 codes with their associated descriptive labels using the Center for Medicare and Medicaid Services label database. In the third step, we stratified by match type (1-to-1, eg, 1 ICD-9 code to 1 ICD-10 code; 1-to-many, eg, 1 ICD-9 code to multiple ICD-10 codes; or multiple:1, eg, multiple ICD-9 codes to 1 ICD-10 code); and in the fourth step by match precision as determined by the GEM database (eg, exact match, or approximate match). During the fifth and sixth steps, we manually obtained and compared 2 sets of codes with descriptive labels to determine the concordance between the 2 groups. In the seventh step, we removed the duplicate codes, and in the eighth step, performed a backward validation exercise to determine whether the ICD-10 codes derived from the algorithm identified any clinically relevant ICD-9 codes not present in the list used during step 1.

Definition of Successfully Matched Codes

We defined ICD-10 codes as successfully validated if the labels were in concordance with the initial ICD-9 code used for translation upon manual review by two physicians. Our primary deliverable was the percent of valid ICD-10 codes as a proportion of the total ICD-10 codes obtained during the translation process. As our primary deliverable was based on a manual comparison of 2 lists of codes, we calculated no P values.

Step 1: Initial List of ICD-9 Codes With ICD-9 Labels

We used a series of prior publications to identify an initial list of ICD-9 codes to use for translation. Specifically, this encompassed ICD-9 codes used in outcome analysis after carotid revascularization (carotid endarterectomy and carotid artery stenting) and endovascular aortic aneurysm repair (Tables I and II in the Data Supplement). The ICD-9 codes evaluated represented several important outcomes to cardiovascular research, including stroke, heart attack, dysrhythmia, heart failure, and procedural wound infections, among others.

We defined an ICD-9 code as the alphanumeric ICD-9 or ICD-10 designation as determined by the Center of Medicare and Medicaid Services. We defined a label as the text description associated with each respective code as designated by the Center of Medicare and Medicaid Services. All translation between ICD-9 and ICD-10 was performed using Microsoft Excel version 15 using the VLOOKUP and IFERROR commands (Microsoft, Redmond, WA).

Step 2: Translation of ICD-9 to ICD-10 Codes

Using our initial list of ICD-9 codes and labels, we performed forward (ICD-9 to ICD-10) GEM using the publicly available database created by the Center for Medicare and Medicaid Services Transitioned from ICD-9 to ICD-10 Transitioned to the ICD-10 list of billing codes. This transition expanded the number of available diagnosis codes from ≈14000 in ICD-9 to over 69000 with ICD-10. This change created several challenges. First, validated lists of ICD-9 codes previously used to identify specific outcomes can no longer be used with ICD-10 data. Second, there is infrequently an exact match between an individual ICD-9 code and an individual ICD-10 code. Computer-based algorithms have been designed to translate ICD-9 into ICD-10 codes, such as the General Equivalence Mapping (GEM) database, but these resources have not been tested for the purpose of clinical research.

Therefore, it was our objective to use publicly available sources to validate a crosswalk by which a list of known ICD-9 codes could be translated into a list of ICD-10 codes. We studied codes used extensively in our research program as part of a longitudinal effort in cardiovascular outcomes research, with the belief that any lessons learned would apply more broadly to ICD-9 to ICD-10 transition efforts.
Columbo et al; ICD-9 to ICD-10 Diagnosis Code Translation

Medicaid Services and the Centers for Disease Control and Prevention. The GEM database provides mappings to translate between ICD-9 and ICD-10 billing codes. Complete information is available at the National Bureau of Economic Research website (www.nber.org).

Steps 3 and 4: Match Type and Match Precision

We then segregated the ICD-9 codes based on the types of matches returned from the GEM database (step 3). Forward mapping resulted in 4 possible types of matches:

- **1:1**, where a single ICD-9 code mapped to a single ICD-10 code (eg, ICD-9 42741: ventricular fibrillation, ICD-10 I4901: ventricular fibrillation)
- **1:Multiple**, where a single ICD-9 code mapped to multiple ICD-10 codes (eg, ICD-9 34201: flaccid hemiplegia affecting dominant side, to ICD-10 G8101: flaccid hemiplegia affecting right dominant side, and ICD-10 G8102: flaccid hemiplegia affecting left dominant side)
- **Multiple:1**, where multiple ICD-9 codes mapped to a single ICD-10 code (eg, ICD-9 4280: congestive heart failure unspecified, and ICD-9 4289: unspecified heart failure, to ICD-10 L509: heart failure, unspecified)

Figure 1. Translation and validation of International Classification of Diseases, Ninth Revision (ICD-9) codes. Methodological steps. ICD-10 indicates ICD, Tenth Revision.

Circ Cardiovasc Qual Outcomes. 2018;11:e004782. DOI: 10.1161/CIRCOUTCOMES.118.004782 October 2018
• 1: none, where a single ICD-9 code had no corresponding ICD-10 code (eg, ICD-9 9985: postoperative infection not elsewhere classified, no corresponding ICD-10 code identified).

We then segregated codes by match precision as designated by the GEM database (step 4). The database designates an exact or approximate match label based on the match of the respective ICD-9 and ICD-10 codes.

Step 5: Obtain Associated Labels for Translated ICD-10 Codes
As the GEM database does not contain labels for the translated ICD-10 codes, we linked the returned ICD-10 codes with their respective labels. To do this we used the publicly available list of ICD-10 labels located on the Center for Medicare and Medicaid Services website (www.cms.gov). This created a matched set of ICD-10 codes with labels from our initial set of ICD-9 codes with labels.

Step 6: Manual Comparison of ICD-9 to ICD-10 Labels
We next performed a manual comparison between the ICD-9 and matched ICD-10 code labels. We classified codes into concordant (eg, ICD-9 4270: paroxysmal supraventricular tachycardia to ICD-10 L471: supraventricular tachycardia) or discordant (eg, ICD-9 43885: vertigo as late effect of cerebrovascular disease to ICD-10 L69998: other sequelae after unspecified cerebrovascular disease). Label concordance was adjudicated by 2 physician reviewers (Drs Columbo and Kang). We then calculated the percent of concordant versus discordant codes found for each type of match. The list of concordant ICD-10 codes represented the list of validated ICD-10 codes for use.

Steps 7 and 8: Removal of Duplicate ICD-10 Codes, and Backward Translation
Finally, we performed code translation in reverse (ICD-10 to ICD-9). We did this for 2 distinct reasons: first, to determine the percent of the original list of codes which would be returned, and second, to ensure that no clinically relevant codes were obtained that should have been included in our initial list of ICD-9 codes. We first removed any duplicates found within the list of ICD-10 codes (step 7). Then, using the final list of ICD-10 codes, we used the GEM database to backward translate the ICD-10 codes into ICD-9 codes (step 8). We then calculated the percent of the initial list of ICD-9 codes that were obtained, and the number of clinically relevant codes not present in the initial list in step 1.

RESULTS
Mapping Codes for Complications After Carotid Revascularization
We began with 126 ICD-9 diagnosis codes used in our prior work in describing the complications after carotid revascularization (Figure 2). Translation to ICD-10 codes using the GEM database yielded 167 ICD-10 codes (Table III in the Data Supplement). We found 1:1 matches, 1:multiple matches, and multiple:1 matches between the codes. We then used the GEM designation of match precision, categorized as exact or approximate, to further segregate the codes. Ratio of 1:1 matches had both exact and approximate types of precision, while 1:multiple matches returned only approximate ICD-10 codes. Additional details are available in Appendix in the Data Supplement.

ICD-9 Codes for Complications After Carotid Revascularization Which Did Not Match to an ICD-10 Code
There were 13 ICD-9 codes which did not match to any ICD-10 codes (Table IV in the Data Supplement). These frequently represented codes which overlapped with other ICD-9 codes in our list and are further delineated in Appendix in the Data Supplement.

Manual Comparison of ICD-9 and ICD-10 Codes for Carotid Complications
Manual comparison of the ICD-9 and ICD-10 code labels demonstrated 97.6% concordance, with 98.8% agreement between the clinical reviewer evaluations (Table VI in the Data Supplement). New codes found on backward translation are noted in Table VII in the Data Supplement.

After removal of duplicate ICD-10 codes and discordant codes, translation yielded 143 validated ICD-10 codes associated with complications after carotid revascularization, a 14% increase from the initial 126 codes.

Mapping Codes for Complications After Endovascular Aortic Aneurysm Repair
In a fashion similar to process used to review the codes examining complications with carotid revascularization, we next studied the ICD-9 codes we used to identify complications after endovascular aortic aneurysm repair. We began with 97 ICD-9 diagnosis codes (Figure 3). Translation to ICD-10 codes using the GEM database yielded 120 ICD-10 codes (Table V in the Data Supplement). As with the carotid revascularization codes, there were 1:1, 1:multiple, and multiple:1 match types. As previously found for carotid revascularization, 1:1 matches had both exact and approximate types of precision, whereas 1:multiple matches returned only approximate ICD-10 codes. Additional details are available in Appendix in the Data Supplement.
ICD-9 Codes for Complications After Endovascular Aortic Aneurysm Repair That Did Not Match to an ICD-10 Code

There were 18 ICD-9 codes which did not match to any ICD-10 codes. Again, these frequently represented codes which overlapped with other ICD-9 codes in our list and are further delineated in Appendix in the Data Supplement.

Manual Comparison of ICD-9 and ICD-10 Codes for Endovascular Aortic Aneurysm Repair

Manual comparison of the ICD-9 and ICD-10 code labels again demonstrated excellent concordance and agreement (98% and 100%, respectively; Table VI in the Data Supplement). New codes found on backward translation are noted in Table 7 in the Data Supplement.

After removal of duplicate ICD-10 codes and discordant codes, translation yielded 108 validated ICD-10 codes associated with complications after carotid revascularization, an 11% increase from the initial 97 codes.

DISCUSSION

Publicly available sources allowed us to derive a clinically relevant list of ICD-10 codes from a known list of ICD-9 codes used by our group and others to identify complications after carotid revascularization and endovascular aortic aneurysm repair. While the translated ICD-10 codes were generally accurate, we found that...
a subset of codes was discordant upon manual review. Between 10% and 19% of ICD-9 codes had no corresponding ICD-10 code, and the list of resultant ICD-10 codes derived from the mapping exercise was between 11% and 14% larger than the original list of ICD-9 codes. Researchers seeking to use electronic sources for translation of ICD-9 codes to ICD-10 should also expect that errors will occur in 2% to 3% of cases and will require manual attention for identification and resolution. The transition to ICD-10 in October 2015 has made the ICD-9 code lists previously derived obsolete. For researchers to continue to use contemporary claims data for knowledge generation, new lists of ICD-10 codes that permit outcome identification must be created. There are several possible ways to accomplish this goal. First, researchers may create new sets of ICD-10 codes without assistance from prior validated ICD-9 lists. This method is time and resource intensive if not impractical given the nearly 5-fold increase in ICD-10 codes. Ultimately, such methodology would require manually reviewing 69000 ICD-10 diagnosis codes to identify those relevant for use in clinical research. In addition to being labor intensive, this method provides no repeatable framework that other investigators may follow to create lists of ICD-10 codes for event detection. Researchers may wish to use the 8-step framework we have provided to generate ICD-10 codes derived from existing lists of ICD-9 codes. Using duplicative manual comparison, we found that this method yielded clinically valid codes >97% of the time for the two ICD-9 lists we analyzed. Furthermore, while these two lists of codes were obtained from prior longitudinal

Figure 3. Translation and validation of International Classification of Diseases, Ninth Revision (ICD-9) codes.
Endovascular aneurysm repair. ICD-10 indicates ICD, Tenth Revision.
work surrounding outcome identification after two cardiovascular procedures, these events comprised more than just procedure-specific events. The codes reviewed herein encompass an array of multisystem complications which may manifest after diverse medical and surgical interventions. For example, we performed code crosswalks for events, including cardiac dysrhythmias, heart failure, stroke, and ventilator-assisted pneumonia. Therefore, the scope of this work likely extends beyond carotid revascularization and endovascular aortic aneurysm repair and has implications for researchers seeking to create and analyze valid sets of ICD-10 codes for research in many clinical fields.

Interestingly, several ICD-9 codes did not match to any ICD-10 diagnosis codes. This most often occurred when ICD-10 codes were more specific than the starting ICD-9 code. In addition, our initial list of ICD-9 codes used for translation was designed to be sensitive in event detection. For this reason, some ICD-9 codes that were less likely to represent complications but that were close in numerical value to other more important diagnosis codes were included in our event detection algorithm. How these factors might affect event detection in claims data, and whether these events represent true clinical outcomes is an area of active investigation for our group.

ICD-10 is likely to be in widespread use for many years. The enhanced granularity of ICD-10 offers many new opportunities for research advances, but its utilization will require the creation of new sets of billing codes to accurately represent clinical events. The derivation and validation of these billing codes is not the responsibility of any one investigator, nor should investigators be working in isolation to accomplish this task. With these things in mind, The Dartmouth Institute has created a public website, where advances in ICD-10 billing code derivation may be stored and easily accessed by any investigator from any institution (http://www.dartmouthatlas.org/pages/ICD10). We implore investigators also navigating the transition from ICD-9 to ICD-10 to collaborate with the many investigators across the country seeking similar goals.

Our method has several limitations. First, the GEM database was designed for billing rather than clinical research, and as such has inherent limitations in ICD-10 code generation. Second, while we took care to exhaustively review our coding algorithms, it remains possible that other relevant codes exist which were not identified with this method. Finally, while our sets of ICD-9 codes represented complications involving multiple organ systems, we cannot comment on the validity of ICD-10 code generation beyond the events that we studied. Nevertheless, this methodological framework provides a starting point for investigators seeking to derive sets of ICD-10 codes from their existing lists of ICD-9 codes.

CONCLUSIONS

We found that by using publicly available files, we were able to translate previously validated sets of ICD-9 codes used in clinical research for two different cardiovascular procedures into valid ICD-10 codes suitable for event identification in claims-based data. We found that ICD-10 coding lists were likely to be 11% to 14% larger than their ICD-9 counterparts and that 2% to 3% of codes generated by automated matching systems would not be deemed useful during a clinically rigorous evaluation. Cardiovascular researchers need to emphasize a systematic approach in developing and validating ICD-10 coding algorithms to ensure accurate assessment of outcomes.

REFERENCES

Columbo et al; ICD-9 to ICD-10 Diagnosis Code Translation


