Quality Assurance of Demographics Consistency between Veterans Affairs and Medicare Data

Aize Cao[1,2], Margaret Gonsoulin[3], Kristin de Groot[3], Elizabeth Hanchrow[1,2], Daniel Park[1,2], Kristine Lynch[4,5], Denise Hynes[3], Scott L. DuVall[4,5], Michael E. Matheny[1,2], Stephen A. Deppen[1,2]

[1]Tennessee Valley Healthcare System, Veterans Affairs Medical Center, Nashville, TN; [2]Vanderbilt University, Nashville, TN
[3]VA Information Resource Center, Edward Hines Jr. VA Hospital, Hines, IL
[4]VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, UT; [5]University of Utah, Salt Lake City, UT

INTRODUCTION

VINCI (VA Informatics and Computing Infrastructure) began transforming VA Clinical Data Warehouse (CDW) to the OMOP CDM in 2015. Starting in 2017, VINCI collaborated with VIReC (VA Information Resource Center) to transform VA Medicare data into the OMOP CDM. Demographic information in electronic medical records and claims data were collected from multiple sources at multiple time points. Longitudinal data commonly have conflicting states of what should be relatively stable demographic variables, namely date of birth (DOB), gender and race. Thus, demographic data often require a great deal of cleaning, costing researchers' significant time and effort.

Between these two large datasets we had an opportunity to compare and validate each demographic variable and estimate the potential reduction in missing or conflicting data.

METHODS

The VA cohort included over 23 million Veterans who accessed the VA healthcare system between January 2000 and December 2016. Before converting the VA CDW demographic information into the OMOP person table, a multi-faceted effort was taken to evaluate conflicting data and determine the unique demographic values for each person based on the best practice logic developed by content experts and source data managers in VIReC and VINCI.[1,2,3,4]

The Medicare data comprised 12 million Veterans who have ever been enrolled in Medicare. OMOP-Medicare DOB and gender were sourced from the Medicare vital status file. Due to underreported minority groups[5], OMOP-Medicare race was obtained from two sources, imputed race (RTI Race) in beneficiary summary file and the race from the Medicare vital status file.

RESULTS

Table 1. Consistency of gender in OMOP-VA and OMOP-Medicare person tables

		Gender in OMOP-CMS person table			
		Male	Female	Unknown	Grand Total
Gender in OMOP-VA person table	Male	6,246,548	33,366	2	6,279,916
	Female	40,132	627,399	1	667,532
	Unknown	3,906	1,273	0	5,179
	Grand Total	6,290,586	662,038	3	6,952,627

The two OMOP CDMs were linked through scrambled SSN resulting in about 7 million matched Veterans. We found the concordance for DOB between the two OMOP CDMs (n=6,523,073, 93.82%). No missing DOB was observed in the OMOP-Medicare cohort, but 37,537 Veterans were missing DOB in the OMOP-VA cohort (0.54%).

Gender matched 98.87% across the two datasets (**Table 1**). Three patients in OMOP-Medicare had missing gender and a non-missing gender value in OMOP-VA. Over 5,000 patients with unknown gender in OMOP-VA had non-unknown value in OMOP-Medicare.

Unknown race was the most common result in the OMOP-VA dataset (33.2%). Unknown race may arise if the individual refuses to identify their race, if race is missing or if conflicting race values could not be resolved. This high rate of unknown race caused the lack of consistency (36.7%) observed for race between the two datasets (**Table 2**). About 2.3 million patients with missing or unknown race in OMOP-VA were not missing race in OMOP-Medicare. Specifically, almost 1.9 million patients missing race in OMOP-VA were identified as white in OMOP-Medicare data, half of those identified as white in the VA dataset. On the other side, 227,934 patients with unknown race in OMOP-Medicare could be found in OMOP-VA, assuming the unknown race in opposite database are accurate.

These two sources put together could be used to assign a race to all but 120,314 members of the cohort, assuming resolution of 130,591 conflicting race values between the two datasets. In this practice, the unknown race in OMOP-VA were re-identified by VA CDW data, which is also an approach of quality assurance for the OMOP-VA race logic during the source data extraction, cleaning, and transformation.

Table 2. Consistency of race in OMOP-VA and OMOP-Medicare person tables

		Race in OMOP-CMS person table					
		White	Black or African American	Asian	American Indian or Alaska Native	Unknown	Grand Total
Race in OMOP-VA person table	White	3,744,379	22,807	6,775	9,897	195,855	3,980,073
	Black or African American	19,536	505,316	509	512	12,923	538,796
	Asian	37,894	642	17,883	232	11,154	67,805
	American Indian or Alaska Native	25,822	2,878	3,087	14,315	8,002	54,104
	Unknown	1,943,052	211,682	24,911	11,890	120,314	2,311,849
	Grand Total	5,771,043	743,325	53,165	36,846	348,248	6,952,627

CONCLUSION

This study found high consistency for gender and DOB between two OMOP CDMs. Low consistency for race occurred due to the high portion of unknown race in VA data. Race data from Medicare could reduce the missing VA race data by 33.2%. We showed a potential improvement in demographic data accuracy above existing cleaning algorithms by comparing and combining the OMOP-VA and OMOP-Medicare datasets.

ACKNOWLEDGEMENTS

This study's work supported with resources and the use of facilities at the TVHS and Salt Lake City VA, and is funded by VA HSR&D VINCI (Project number RES 13-457). Support for VA/Medicare Data provided by VA HSR&D, VA Information Resource Center (Project numbers SDR 02-237 and 98-004).

REFERENCES

1. Kevin T. Stroupe, Elizabeth Tarlov, Qiuying Zhang, Thomas Haywood, Arika Owens, Denise M. Hynes, Use of Medicare and DOD data for improving VA race data quality, Journal of Rehabilitation Research & Development, Vol 47 (8), Page 781-796, 2010 2.
2. Data quality analysis team, CDW race data and multiple races, http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/CDW_Race_Data_and_Multiple_Races.p df
3. Gonsoulin, Margaret. Using SQL to "Sort Out" Race in CDW: A method for cleaning multiple values of race. The Researcher's Notebook; no. 6. Hines, IL: VA Information Resource Center; 2016. http://vaww.virec.research.va.gov/Notebook/RNB/RNB6-CDW-SQL-to-Sort-Out-Race-CY16.pdf
4. Best practices guide race data, http://vaww.vhadataportal.med.va.gov/Portals/0/DataQualityProgram/Reports/Best_Practices_Guide_Race_Data.pdf
5. Eicheldinger, C. Bonito, A. More Accurate Racial and Ethnic Codes for Medicare Administrative Data. Health Care Financing Review, Spring 2008, Volume 29, Number 3. https://www.cms.gov/Research-Statistics-Data-andSystems/Research/HealthCareFinancingReview/downloads/08Springpg27.pdf

VA Informatics and Computing Infrastructure