Evaluating the Role of Past Clinical Information on Risk Adjustment

John K. Lin, MD,*† Juliette Hong, MS,‡§ Ciaran Phibbs, PhD,‡|| Peter Almenoff, MD,|| and Todd Wagner, PhD‡##**

Objective: The objective of this study was to evaluate whether incorporating historical clinical information beyond 1 year improves risk adjustment.

Data Sources: Administrative data from the Department of Veterans Affairs and Medicare (for veterans concurrently enrolled in Medicare) for fiscal years (FYs) 2011–2015.

Study Design: We regressed total annual costs on Medicare hierarchical condition category indicators and risk scores for FY 2015 in both a concurrent and a prospective model using 5-fold cross-validation. Regressions were repeated incorporating clinical information from FY 2011 to 2015. Model fit was appraised using R^2 and mean squared predictive error (MSPE).

Data Collection: All veterans affairs users ($n=3,254,783$) with diagnostic information FY 2011–2015.

Principal Findings: In a concurrent model, adding additional years of historical clinical information (FY 2011–2014) did not result in substantive gains in fit (R^2 from 0.671 to 0.673) or predictive capability (MSPE from 1956 to 1950). In a prospective model, adding additional years of historical clinical information also did not result in substantive gains in fit (R^2 from 0.334 to 0.344) or predictive capability (MSPE from 3988 to 3940).

Conclusion: Incorporating historical clinical information yielded no material gain in risk adjustment fit.

Key Words: risk adjustment, cost, performance measurement, health economics

(*Med Care* 2020;58: 169–173)

Risk adjustment is a foundational component of health services research. Health administrative databases are used to track health system performance in the form of outcomes and costs. However, accurate evaluations of health system performance require adjusting for patients' underlying clinical risk.

Numerous risk adjustment algorithms have been developed over the past 2 decades.¹ So far, these algorithms have incorporated clinical information from the present year to create a risk score for the same or following year. Risk adjustment typically relies on the International Classification of Diseases (ICD) diagnostic codes to classify patients into clinical categories. Earlier studies have demonstrated that incorporating prior years of clinical information may modestly improve prediction of clinical endpoints, such as mortality.^{2–6} There are 2 ways this may occur: (1) clinical severity is related to time with disease, and incorporating historical clinical information may serve as a proxy for severity; (2) clinical memory can also partially correct for missing data, which could be due to out-of-plan use or poor coding. However, these studies evaluated risk adjustment using older tools, such as the Charlson Comorbidity Index, or non-US risk adjustment scores, such as the Multipurpose Australian Comorbidity Scoring System. It remains unknown whether incorporating prior clinical information improves risk prediction using more modern risk adjustment tools, in a United States setting.

The Veterans Affairs (VA) has developed a risk adjustment model (Nosos) that incorporates Medicare Advantage's V21 risk score with additional mental health and pharmacy information to predict health care costs. In a validation study, it performed on par with the DxCG commercial software.⁷ We sought to evaluate whether incorporating clinical information from the previous 4 years improves the risk score's predictive capability.

METHODS

Data

We analyzed administrative data from the Department of Veterans Affairs for fiscal years (FYs) 2011 to 2015. Utilization and diagnostic data were taken from national databases (National Patient Care Database and Patient Treatment Files). Cost data

From the *University of Pennsylvania Perelman School of Medicine; †Leonard Davis Institute, University of Pennsylvania, Philadelphia, PA; ‡Health Economics Resource Center (HERC), VA Palo Alto Health Care System, Palo Alto; §Department of Anesthesiology, Perioperative and Pain Medicine; ||Department of Pediatrics, Stanford University; ¶Operational Analytics and Reporting, Office of Informatics and Analytics; #Center for Innovation to Implementation, VA Palo Alto Health Care System, Palo Alto; and **Department of Health Research and Policy, Stanford University, Stanford, CA.

P.A. is now retired from the Department of Veterans Affairs.

The authors declare no conflict of interest.

Reprints: John K. Lin, MD, 3400 Civic Center Boulevard, 10th Floor Perelman Center South Pavilion, Philadelphia, PA 19104. E-mail: john.lin@pennmedicine.upenn.edu.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.lww-medicalcare.com.

Written work prepared by employees of the Federal Government as part of their official duties is, under the U.S. Copyright Act, a "work of the United States Government" for which copyright protection under Title 17 of the United States Code is not available. As such, copyright does not extend to the contributions of employees of the Federal Government. ISSN: 0025-7079/20/5802-0169

were obtained from the Managerial Cost Accounting and the Health Economic Research Center Average Cost Databases. For the fraction of VA patients who also used Medicare, we also incorporated Medicare costs and diagnostic data. We included patients where hierarchical condition category (HCC) information was not missing between 2011 and 2015. The final analytical dataset had 3,254,783 VA users.

Describing the Original Nosos Model

The Nosos risk adjustment model incorporates a patient's medical history in the form of the Centers for Medicare and Medicaid Services (CMS), V21 HCC risk score (institutional weights if a patient stayed in any institution for that FY, otherwise community weights). The HCC risk score is composed of the individual's demographics and ICD codes.⁸ In addition to the HCC risk score, Nosos also incorporates demographic covariates, mental health and pharmacy information, and health insurance status. Additional veteran-specific information is included, such as priority-level status, which designates the priority to which medical benefits and resources are allocated to the patient, and whether the patient was in a registry, such as exposure to Agent Orange. Finally, we also incorporated a previously validated frailty index score, which is not in the main Nosos model.⁹ None of these additional measures incorporates clinical information from previous years. The risk adjustment model regresses the square root of total annual VA costs on the covariates above.

Analytic Overview

We compared the current Nosos risk adjustment model against a model incorporating the same variables in addition to clinical information from FYs 2011 through 2015. Our first model was most comparable to Nosos in its current form, where medical information is aggregated as an HCC risk score. Here, we incorporated the patient's aggregated risk scores each year from 2011 to 2015 as additional covariates into the regression, effectively resulting in 4 additional HCC risk score covariates (2011–2014). Although incorporating clinical information as an aggregated HCC score, more detailed clinical information is lost. We therefore also created a second model, which uses each individual HCC indicator from FYs 2011 to 2015 (vs. 2015 alone), resulting in a 4-fold increase in HCC indicator covariates.

We created a third set of models that aggregated HCC covariates from 2011 to 2015 into either 1 set of indicators or 1 score. If a patient had an HCC indicator coded in any year from 2011 to 2015, the patient was considered to have that aggregated HCC indicator. An aggregated HCC score was recalculated based on the CMS, V21 formula from the aggregated set of HCC indicators.

We evaluated the role of adding past clinical information in both concurrent and prospective models. The concurrent model uses a given FY's (2015, or in the case of the clinical memory models, 2011–2015) clinical information to predict costs for the same year (2015). Its purpose is explanatory and can be used to measure health system performance. The prospective model uses a given FY's clinical information (2014, or in the case of the clinical memory models, 2011–2014) to predict costs for the following year (2015). Its purpose is predictive and can be used for allocating future payments.

To assess overfitting, we used 5-fold cross-validation to compare models.¹⁰ We compared each model in its ability to predict current and prospective costs, using the mean squared predictive error (MSPE) and R^2 , averaged across the 5 iterations.

Incorporating clinical memory is hypothesized to improve risk adjustment in part because it can partially correct for missing data; in VA, poor coding is possible because payments are not linked to coding. Some conditions (such as rheumatologic conditions) may also present diagnostic challenges and be difficult to code. To quantify the prevalence of coding gaps among chronic conditions, we tracked conditions listed in the Chronic Conditions Warehouse¹¹ in addition to HCC categories that 2 physicians (J.K.L. and Omar A. Usman) deemed as chronic: chronic pancreatitis (HCC34), inflammatory bowel disease (HCC35), disorders of immunity (HCC47), coagulation defects (HCC48), amputations (HCC173 and HCC189), and major organ transplant or replacement status (HCC186). For the identified chronic conditions, we report the fraction of patients where the diagnostic code was not used in an intermediate year (eg, if it was coded in FYs 2011 and 2014 but not 2012 or 2013). Stages of chronic kidney disease are along a continuum and were aggregated together for this specific purpose. We created another series of regressions to examine whether filling in these coding gaps would improve model fit.

Finally, we evaluated whether additional diagnostic history improves prediction in sparse regressions. Here, we only included demographics (age, sex, marital status, and race) and HCC information.

RESULTS

Descriptive Statistics

Tracking chronic conditions over time identified gaps in diagnostic coding ranging from 3.0% to 23.8%. This included dialysis status (3.0%), human immunodeficiency virus or acquired immunodeficiency syndrome (4.7%), metastatic cancer and acute leukemia (7.1%), Parkinson's and Huntington's diseases (8.9%), schizophrenia (11.4%), chronic obstructive pulmonary disease (19.5%), and chronic hepatitis (23.8%) (Fig. 1).

An additional table demonstrating how HCC indicators and scores change when using 2015 information versus aggregating 2011–2015 information together is included in the Supplementary (Webtable 1, Supplemental Digital Content 1, <http://links.lww.com/MLR/B898>).

In a concurrent model, the risk adjustment model incorporating HCC risk scores from FY 2015 resulted in an R^2 of 0.671 and MSPE of 1956. Incorporating HCC risk scores from FYs 2011 to 2014 to the regression slightly increased R^2 (0.673) and decreased MSPE to 1950. Regressions incorporating individual HCC indicators performed slightly better than regressions using the aggregate HCC risk score. In a regression using individual HCC indicators from FY 2015, the R^2 was 0.692 and MSPE was 1832. By adding individual HCC indicators from FYs 2011 to 2014 to the regression, R^2 increased slightly from 0.695 and MSPE decreased to 1812. Filling the coding gaps identified above as lagged variables resulted in minimal improvement to model performance: R^2 increased to 0.696 and MSPE decreased to 1803 in a regression incorporating lagged indicators (Table 1).

FIGURE 1. Coding gaps from chronic conditions. For each of 52 chronic conditions, the fraction of patients where that chronic condition's diagnostic code was not used in an intermediate year is shown (eg, if it was coded in fiscal years 2011 and 2014 but not 2012 or 2013). AIDS indicates acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

In a prospective model, the risk adjustment model incorporating HCC risk scores from FY 2015 resulted in worse fits and larger prediction error than the concurrent model: an R^2

of 0.334 and MSPE of 3988. Incorporating HCC risk scores from FYs 2011 to 2014 to the regression increased R^2 (0.344) and slightly decreased MSPE to 3940. For the prospective

TABLE 1. Model Fit Statistics From Regressions With and Without Clinical Memory

Models	R^2	MSPE
Concurrent model (predicting 2015 costs)		
HCC risk score		
Nosos* (HCC score, 2015)	0.671	1956
Nosos* (HCC scores, 2011–2015)	0.673	1950
Nosos* (aggregated HCC score, comprised of 2011–2015 data) [†]	0.643	2123
Sparse model: HCC 2015 and demographics only [‡]	0.462	3197
Sparse model: HCC 2011–2015 and demographics only [‡]	0.463	3189
Individual HCC risk indicator		
Nosos* (HCC indicators, 2015)	0.692	1832
Nosos* (HCC indicators, 2011–2015)	0.695	1812
Nosos* (HCC indicators, 2011–2015) with coding gaps filled	0.696	1803
Nosos* (aggregated HCC indicators, comprised of 2011–2015 data) [†]	0.660	2019
Sparse model: HCC indicator 2015 and demographics only [‡]	0.517	2865
Sparse model: HCC indicator 2011–2015 and demographics only [‡]	0.526	2821
Prospective model (predicting 2015 costs)		
HCC risk score		
Nosos* (HCC score, 2014)	0.334	3988
Nosos* (HCC scores, 2011–2014)	0.344	3940
Nosos* (aggregated HCC score, comprised of 2011–2014 data) [†]	0.336	3973
Sparse model: HCC 2014 and demographics only [‡]	0.222	4618
Sparse model: HCC 2011–2014 and demographics only [‡]	0.248	4469
Individual HCC risk indicator		
Nosos* (HCC indicators, 2014)	0.344	3940
Nosos* (HCC indicators, 2011–2014)	0.356	3864
Nosos* (HCC indicators, 2011–2014) with coding gaps filled	0.359	3833
Nosos* (aggregated HCC indicators, comprised of 2011–2014 data) [†]	0.352	3895
Sparse model: HCC indicator 2014 and demographics only [‡]	0.256	4420
Sparse model: HCC indicator 2011–2014 and demographics only [‡]	0.279	4287

*In addition to HCC information, the Nosos models incorporate mental health and pharmacy information, JEN frailty index score as well as health insurance status. Additional veteran-specific information is also included, such as priority-level status, which designates the priority to which medical benefits and resources are allocated to the patient, and whether the patient was in a designated registry, such as exposure to Agent Orange.

[†]In the aggregated HCC models, multiple years of information were used to create indicators (eg, if the HCC indicator for HIV was present in any year from 2011 to 2015, there was a single indicator variable for HIV that was considered present) or a single score.

[‡]The sparse models include age, sex, marital status, race and HCC information. They do not incorporate the additional variables specified above.

HCC indicates hierarchical condition category; HIV, human immunodeficiency virus; MSPE, mean squared predictive error.

model, regressions incorporating individual HCC indicators performed slightly better than regressions using the aggregate HCC risk score. In a regression using individual HCC indicators from FY 2015, the R^2 was 0.344 and MSPE was 3940. By adding individual HCC indicators from FYs 2011 to 2014 to the regression, R^2 increased slightly to 0.356. MSPE decreased to 3864. Filling the coding gaps identified above as lagged variables resulted in minimal improvement to model performance: R^2 increased to 0.359 and MSPE decreased to 3833 in a regression incorporating lagged indicators (Table 1).

Sparse regressions that only included demographic information (age, sex, marital status, and race) and HCC diagnostic information performed more poorly than the Nosos models. Incorporating past diagnostic information resulted in improvements in the prediction that were slightly larger than those seen in the Nosos models, but were still modest (Table 1).

Regressions using an HCC score or set of HCC indicators that aggregated diagnostic information over multiple years performed more poorly than regressions only incorporating 1 year of HCC information (Table 1).

DISCUSSION

Surprisingly, incorporating clinical information from previous years did not substantially improve risk adjustment performance as measured by R^2 and MSPE in both concurrent and prospective models.

Our results contrast with those of prior studies, which showed that for older risk adjustment tools (such as the Charlson Comorbidity Index), prior years of clinical information modestly improved clinical outcome prediction.^{2–6} One possibility is that past clinical information is less useful for predicting costs than clinical outcomes.

It is also possible that for newer risk adjustment tools (such as Medicare HCC's) that more accurately predict outcomes, the marginal gain of past clinical information is small. For instance, the gains with adding prior diagnostic information in a sparse model (HCC plus demographics) were larger than those in the richer Nosos model.

One last possibility is that heterogeneity in disease progression may render historical ICD data a suboptimal predictor of both costs and outcomes. For instance, in diseases such as cancer, if the patient has progressive metastatic disease, time since diagnosis may portend high costs. Conversely, if the cancer is caught early on and treated with definitive therapy, or if the patient chooses to enroll in hospice, time since diagnosis may also predict low costs. Adequacy of disease control is the key mediator, which is unobserved in administrative data.

Levels of coding discontinuities were notable for a number of conditions (eg, chronic hepatitis, chronic obstructive pulmonary disease, and vascular disease). These gaps were anticipated, as VA is capitated and clinicians are not paid based on coding, in contrast to Medicare Advantage where intensive coding is rewarded.¹² What was surprising was despite correcting the discontinuities, the risk adjustment model performance did not substantively improve. This observation suggests it is possible that in the VA, coding fidelity may be high when considered as a measure of utilization or cost (even if well-managed or latent conditions that contribute little to cost may be under-coded). Our study suggests that further improvements in risk adjustment models may require new variables, such as socioeconomic status, markers of disease severity, which may not be presently captured in administrative data.

Limitations

This study was performed in the VA. Major improvements in model fit in a Medicare population seem unlikely,

however, given diagnostic information is necessary for Medicare Fee for Service billing, and it directly relates to payments in Medicare Advantage plans. We used standard rules for assigning HCC's (where 1 diagnosis is sufficient), which results in increased sensitivity and decreased specificity compared with alternative specifications (eg, where > 2 diagnoses would be required to assign an HCC). Such specifications may warrant further study.

CONCLUSION

Incorporating clinical information from previous years as lags did not substantively improve cost-related risk adjustment in a VA population.

ACKNOWLEDGMENT

The authors thank Dr Omar A. Usman for his expertise in reviewing chronic conditions for this manuscript.

REFERENCES

- Schone E, Brown RS. *Risk Adjustment: What Is The Current State Of The Art, And How Can It Be Improved*. Mathematica Policy Research: Research Synthesis Report No. 25. Minneapolis, MN: Robert Wood Johnson Foundation; 2013.
- Preen DB, Holman CD, Spilsbury K, et al. Length of comorbidity lookback period affected regression model performance of administrative health data. *J Clin Epidemiol*. 2006;59:940–946.
- Zhang JX, Iwashyna TJ, Christakis NA. The performance of different lookback periods and sources of information for Charlson comorbidity adjustment in Medicare claims. *Med Care*. 1999;37:1128–1139.
- Chen G, Lix L, Tu K, et al. Influence of using different databases and “look back” intervals to define comorbidity profiles for patients with newly diagnosed hypertension: implications for health services researchers. *PLoS One*. 2016;11:e0162074.
- Dobbins TA, Creighton N, Currow DC, et al. Look back for the Charlson index did not improve risk adjustment of cancer surgical outcomes. *J Clin Epidemiol*. 2015;68:379–386.
- Stukenborg GJ, Wagner DP, Connors AF. Comparison of the performance of two comorbidity measures, with and without information from prior hospitalizations. *Med Care*. 2001;39:727–739.
- Wagner TH, Upadhyay A, Cowgill E, et al. Risk adjustment tools for learning health systems: a comparison of DxCG and CMS-HCC V21. *Health Serv Res*. 2016;51:2002–2019.
- Pope GC, Kautter J, Ellis RP, et al. Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financ Rev*. 2004;25:119–141.
- Kinosian B, Wieland D, Gu X, et al. Validation of the JEN frailty index in the National Long-Term Care Survey community population: Identifying functionally impaired older adults from claims data. *BMC Health Serv Res*. 2018;18:1–12.
- Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer-Verlag; 2001.
- Centers for Medicare and Medicaid Services. Chronic conditions data warehouse: condition categories; 2018. Available at: www.ccwdata.org/web/guest/condition-categories. Accessed January 31, 2019.
- Geruso M, Layton T. Upcoding: evidence from Medicare on squishy risk adjustment. NBER Working Paper #21222. May 2015, Revised April 2018.