UCSF

Title

Development and validation of novel multimorbidity indices for older adults.

Permalink

Journal

Authors

Steinman, Michael

Jing, Bocheng

Shah, Sachin

et al.

Publication Date

DOI

Development and validation of novel multimorbidity indices for older adults

Michael A. Steinman, MD[1,2], **Bocheng Jing, MS**[1,2], **Sachin J. Shah, MD, MPH**[3], **Anael Rizzo, BS**[1,2,4], **Sei J. Lee, MD, MAS**[1,2], **Kenneth E. Covinsky, MD, MPH**[1,2], **Christine S. Ritchie, MD, MSPH**[5], **W. John Boscardin, PhD**[1,2,6]

[1]Division of Geriatrics, UCSF, San Francisco, California, USA

[2]The San Francisco VA Health Care System, San Francisco, California, USA

[3]Division of Hospital Medicine, UCSF, San Francisco, California, USA

[4]David Geffen School of Medicine at UCLA, San Francisco, California, USA

[5]Division of Palliative Care and Geriatric Medicine, Massachusetts General Hospital and the Mongan Institute Center for Aging and Serious Illness, Boston, MA, USA

[6]Department of Epidemiology and Biostatistics, UCSF, San Francisco, California, USA

Abstract

Background: Measuring multimorbidity in claims data is used for risk adjustment and identifying populations at high risk for adverse events. Multimorbidity indices such as Charlson and Elixhauser scores have important limitations. We sought to create a better method of measuring multimorbidity using claims data by incorporating geriatric conditions, markers of disease severity, and disease-disease interactions, and by tailoring measures to different outcomes.

Methods: Health conditions were assessed using Medicare inpatient and outpatient claims from subjects age 67 and older in the Health and Retirement Study. Separate indices were developed for ADL decline, IADL decline, hospitalization, and death, each over 2 years of follow-up. We validated these indices using data from Medicare claims linked to the National Health and Aging Trends Study.

Correspondence: Michael A. Steinman, Division of Geriatrics, UCSF, 4150 Clement Street, Box 181G San Francisco, CA 94121, USA. mike.steinman@ucsf.edu.

Results: The development cohort included 5012 subjects with median age 76 years; 58% were female. Claims-based markers of disease severity and disease-disease interactions yielded minimal gains in predictive power and were not included in the final indices. In the validation cohort, after adjusting for age and sex, c-statistics for the new multimorbidity indices were 0.72 for ADL decline, 0.69 for IADL decline, 0.72 for hospitalization, and 0.77 for death. These c-statistics were 0.02–0.03 higher than c-statistics from Charlson and Elixhauser indices for predicting ADL decline, IADL decline, and hospitalization, and <0.01 higher for death ($p < 0.05$ for each outcome except death), and were similar to those from the CMS-HCC model. On decision curve analysis, the new indices provided minimal benefit compared with legacy approaches. C-statistics for both new and legacy indices varied substantially across derivation and validation cohorts.

Conclusions: A new series of claims-based multimorbidity measures were modestly better at predicting hospitalization and functional decline than several legacy indices, and no better at predicting death. There may be limited opportunity in claims data to measure multimorbidity better than older methods.

Keywords

claims data; functional impairment; multimorbidity; older adults; prognostic models

INTRODUCTION

The presence of multiple chronic conditions has major impacts on treatment complexity, healthcare costs, and clinical outcomes. As such, measures of multimorbidity—which collapse the cumulative impact of multiple chronic conditions into a single score that is easy to use and interpret—are used by health systems, payors, and researchers for risk adjustment and to identify people at elevated risk of poor outcomes.

This has spawned a substantial body of scholarship to develop summative measures of multimorbid burden, particularly using administrative data sources which can facilitate their widespread use.[1–3] Among the first was the Charlson Comorbidity Index, which evaluates for 19 chronic conditions. Each condition that is present is assigned a weight of 1 through 6 based on its contribution to mortality in the original study of hospitalized adults; the sum of these weights yields an individual's overall comorbidity score.[4] Dozens of other indices, most using a variety of Charlson's sum-of-weighted-conditions approach, have subsequently been developed and validated.[4–13] None of this group of indices proved unambiguously better than the others, and all suffer from generally limited predictive power.[3] As a result, the Charlson index and Elixhauser approach (another early effort, specifically tailored to claims data) remain in widespread use for research applications.[7,8,14] Meanwhile, more complex approaches including the CMS-Hierarchical Condition Categories (CMS-HCC) model—which identifies and accounts for the most serious condition a given person has among related conditions—have been developed with healthcare costs in mind and are commonly employed by health systems and payors for risk adjustment.[15]

Several limitations of past approaches suggest opportunity to substantially improve claims-based multimorbidity measurement. First, most multimorbidity indices were developed to predict death, hospitalization, and/or healthcare costs.[8,16] Yet, developing impairments

in activities of daily living is an outcome of great importance for older adults and the conditions that may augur such declines—for example, osteoarthritis—may be different than those that lead to hospitalization or death.[5,17–22] Thus, using different indices to predict different outcomes may be better than a one-size-fits-all approach. Second, clinicians know that the severity of a given disease and the presence of disease-disease interactions can have critical prognostic implications. Yet, while some multimorbidity indices based on patient interviews or chart review incorporate markers of disease severity,[23–26] this has not been extended to indices that use claims data. Similarly, disease-disease interactions have not been incorporated into multimorbidity indices. Finally, multimorbidity indices have largely ignored geriatric syndromes such as falls, pressure ulcers, and weight loss.[13] This is a missed opportunity, as these conditions are vital markers of overall health and have important prognostic implications.[27,28]

We thus sought to develop and validate a new family of claims-based multimorbidity indices, each tuned to a different clinical outcome, that take advantage of these opportunities for improvement. Our goal was to create a new, easy-to-use system for measuring multimorbidity using claims data that performs better than legacy approaches in the applications for which such indices are typically used.

METHODS

Data sources and subjects

Development cohort: We developed new multimorbidity indices using data from the United States' Health and Retirement Study (HRS), a nationally representative biennial survey. We included community-dwelling adults age 67 and older who participated in the 2010 HRS interview, were enrolled in fee-for-service Medicare Parts A and B in the 2 years prior, and consented to linking their Medicare data. For models focused on ADL and IADL outcomes, we excluded subjects who did not complete full ADL and IADL assessments at the baseline and follow-up interviews.

Validation cohorts: Indices were externally validated in National Health and Aging Trends Study (NHATS), another nationally representative survey. The baseline year was 2011 and follow-up year was 2013. Inclusion criteria and approaches were identical to the development cohort. As a secondary validation cohort, we returned to HRS, but used 2014 instead of 2010 as the baseline year (and with otherwise identical inclusion and exclusion criteria; Supplementary Figure S03).

Measures—Predictors

Clinical conditions were ascertained using Medicare Parts A and B data (Inpatient, Carrier, and Outpatient files). Because our indices were designed only to measure multimorbidity and not to serve as comprehensive risk prediction indices, we did not include other factors such as socioeconomic characteristics or history of health services utilization.

Based on a literature review we identified 129 diseases, conditions, and clinical syndromes (hereafter collectively termed "conditions") that could be encoded in claims data and might improve discrimination for a multimorbidity index (see Supplementary Text and Tables

S02 and S04 for details). Conditions were mapped to ICD-9-CM codes using the HCUP Clinical Classification Software (CCS), supplemented by code mapping schemas from other sources (Supplementary Tables S04 and S05).[29–34] We defined a condition as present if the corresponding ICD-9-CM code was recorded in at least one outpatient or inpatient encounter over the past 2 years.[35]

There are no well-established methods for identifying disease severity in claims data that are applicable to a wide range of diagnoses. Thus, based on existing literature and clinical judgment we defined seven potential markers of disease severity for each candidate condition, for example having four or more outpatient encounters for that condition in the past year.[36–40] To identify potential disease-disease interactions, we conducted a binning exercise with four experienced geriatrician researchers, where each panelist was asked to sort these conditions into groups, with the expectation that an older adult with two or more conditions in a group would be at especially high risk of adverse health outcomes due to harmful synergy between the conditions (Supplementary Table S06).[41]

Measures—Outcomes

Death and acute medical hospitalization over 2 years were assessed using Medicare claims data. Decline in six activities of daily living (ADLs) and five instrumental activities of daily living (IADLs) over 2 years were assessed using self- and proxy-report data from HRS and NHATS and was defined as requiring help with a greater number of activities at the 2-year follow-up interview compared to the baseline interview (Supplementary Text S02).

To address death as a competing outcome, we classified subjects who died prior to 2-year follow-up as died. For example, the ADL decline outcome comprised three categories: alive without ADL decline, alive with ADL decline, or dead. We used a similar approach to the competing risk of death for the hospitalization outcome, based on input from our advisory panel who felt it would be clinically meaningful to clinicians and patients.

Analyses

Descriptive analyses accounted for the complex survey design of both HRS and NHATS. Models were developed using normalized weights, as accounting for clustering and stratification created difficulties fitting models with small cell sizes.[42] Testing affirmed that point estimates for c-statistics were identical with or without this adjustment.

We developed models in the derivation cohort using multinomial logistic regression which controlled for age and sex and accounted for the competing risk of death. In these models, we explored modeling age using restricted cubic splines with three knots, but found it provided minimal benefit for model performance compared to modeling age as a simple linear term (Supplementary Text S02), so subsequent analyses used the latter approach. We illustrate our development process using the example of the ADL decline index. First, we assessed each of 129 candidate conditions separately. We retained conditions that were present in 1% of the population, clinically distinct from other candidate conditions, and predictive of ADL decline and/or the competing risk of death at $p < 0.20$. As a sensitivity analysis to guard against premature exclusion of conditions, we later included in our models all conditions which had been excluded at this stage on the basis of no association with

outcomes (i.e., p 0.20). For each outcome, reintroducing these conditions en masse changed the c-statistic of our models by <0.001.

The remaining conditions were entered into a multivariable, multinomial logistic regression model, to which we applied backward selection with a retention threshold of $p < 0.05$ using the PROC LOGISTIC procedure in SAS 9.4. While stepwise regression methods have important limitations in certain settings, their performance for predictive modeling applications (without implications for causal inference) is similar to that of newer methods and have advantages of simplicity and transparency,[43] and for our goals AIC-based methods of variable selection are overly inclusive and BIC-based ones too restrictive.

To transform the resultant multinomial model into an easy-to-use index, we excluded conditions with negative parameter estimates and converted the remaining parameter estimates for the ADL decline outcome into integer scores using methods of Sullivan et al.[44] These integers were summed into a patient-level multimorbidity score, with minimal effect on c-statistics (see Supplementary Table S09). We repeated this process for each of our three other outcomes, starting at the second step to enhance opportunities for consistency of conditions across each of the indices. The index focused on death was modeled using logistic rather than multinomial regression models. Methods for testing disease severity markers and disease-disease interactions are described in Supplementary Text S02.

Validation and comparison with legacy models

To test overfitting, we internally validated the multivariable models on which our indices were based using 1000 bootstrap samples.[45] Next, we externally validated our indices among subjects in an independent validation cohort (NHATS) and compared their performance against claims-based versions of four legacy indices: the Charlson Comorbidity Index (Quan adaptation), Elixhauser Comorbidity Score (van Walraven index adaptation, using Quan ICD-9 codes), Functional Comorbidity Index (Kumar adaptation), and CMS-Hierarchical Condition Categories score (CMS-HCC).[29,46,47] We chose these four because the Charlson and Elixhauser approaches are the most commonly used in research, the FCI has historically been the most well-known index for predicting physical function (although recent work by Wei et al has advanced this area),[10,11] and CMS-HCC is commonly used by health systems and payors. In recent years, there has been increasing use of frailty indices but these are conceptually distinct from indices of multimorbidity.[30,48–50]

Our approach to comparing the new indices to legacy ones was informed by the two predominant use cases for multimorbidity indices.[51,52] First, these indices are commonly used as one variable in multivariable models for risk adjustment and/or outcome prediction. As the relevant metric is a continuous measure of the index's impact on model performance, our outcome of interest was the c-statistic.[53] We calculated c-statistics from models that contained each subject's index score, age, and sex (we included age and sex because they are fundamental in the use cases of interest). Because c-statistics cannot be computed from multinomial models with more than two outcomes, we excluded decedents from these calculations (see Supplementary Text S02).[54] As a secondary outcome we assessed the Polytomous Discrimination Index (PDI), which is analogous to a c-statistic for more than

two outcomes, on the full cohort (which included decedents).[55] To assess calibration, we inspected calibration plots and calculated the Hosmer–Lemeshow statistic.

The second use case for multimorbidity indices is for stratification by degree of multimorbidity, for example in subgroup analyses in an observational study or randomized trial, or by health systems which seek to identify which older adults are at highest risk of adverse future outcomes. We thus compared new and legacy indices using Decision Curve Analysis.[52,56] Models used for this analysis included age, sex, and each index. Based on clinical judgment about likely use cases we set decision thresholds a priori at 5%, 10%, 25%, and 50%. As secondary metrics we also calculated the Integrated Discrimination Index and Net Reclassification Index; these approaches have important limitations but provide complementary information to our primary outcome.[55,57–60]

Alternate approaches and sensitivity analyses

We explored alternate analytic approaches including best subsets selection, Lasso regression, and random forest techniques. None of these techniques yielded a meaningful improvement in c-statistic or simplicity compared with our main approach, consistent with other research which suggests that different methods yield similar performance in prognostic model development.[61]

Reporting of this research followed TRIPOD guidelines (Supplementary Table S01).[62] Additional information on methods can be found in the Supplementary Text and Tables S02–S06. This research was approved by the Institutional Review Boards of the University of California, San Francisco and the San Francisco VA Medical Center. Analyses were performed in SAS version 9.4 and R version 4.0.5. R package "dcurve" was used to plot C-statistics comparisons for novel versus legacy indices and decision curves, package "mcca" was used to calculate NRI and IDI indices, and package "rms" was used for model calibration. Statistical code to apply indices in claims data is available at https://github.com/UCSFGeriatrics/Repository.

Role of funding source

This research was funded by the National Institutes of Health. The funder had no role in study design, conduct, or interpretation, nor any role in the decision to publish the results.

RESULTS

Index development and validation

Index development: The derivation cohort included 5102 subjects. Median age was 76 years and 58.7% were female (Table 1 and Supplementary Tables S07 and S08). At the end of the 2-year follow-up period, 8.7% of subjects were alive and with ADL decline, 10.9% were alive and with IADL decline, 25.5% were alive and had been hospitalized in the past 2 years, and 8.9% had died. The index development process yielded the four indices shown in Table 2. Heart failure and COPD were common to all four indices, whereas other conditions were included in only a few or just one index. Best subsets testing revealed multiple other

sets of conditions with similar predictive power as those included in our final models (data not shown).

Claims-based markers of condition severity—for example, having four or more visits for a given condition in the prior year—did not yield meaningful improvement in model c-statistics (maximal gain in c-statistic 0.007). Similarly, addition of markers of disease-disease interactions—for example, the combination of cognitive impairment with auditory impairment—had minimal to no impact on c-statistics compared to the base approach (Supplementary Text S10).

Internal validation: We internally validated our index in the derivation cohort using 1000 bootstrap samples. Model optimism, which reflects overfitting, was 0.007 for ADL decline, 0.005 for IADL decline, and 0.003 for hospitalization and death, consistent with a minor degree of overfitting (Supplementary Table S11).

External validation: Subject characteristics and outcome rates in the external validation cohort were generally similar to the derivation cohort (Table 1 and Supplementary Tables S07 and S08). For each index, rates of the outcome of interest and the competing risk of death increased across quartiles of score (Figure 1).

After controlling for age and sex, c-statistics in the validation cohort were 0.722 for ADL decline, 0.686 for IADL decline, 0.722 for hospitalization, and 0.773 for death (Table 2). The equivalent PDI values were 0.509 for ADL decline, 0.514 for IADL decline, and 0.523 for hospitalization (Supplementary Table S12). Calibration for each of the four indices was robust (Supplementary Figure S13).

Comparison to legacy models

In the validation cohort, after controlling for age and sex the new indices had c-statistics 0.01–0.03 higher than the Charlson Index, Elixhauser Index, and Functional Comorbidity Index for predicting ADL decline, IADL decline, and hospitalization ($p < 0.05$ for all comparisons except FCI for IADL decline; Figure 2 and Supplementary Table S14). C-statistics for all outcomes were similar to those of the CMS-HCC score (c-statistic difference <0.01, $p > 0.05$ for all comparisons). All the indices except FCI performed similarly in their ability to predict death (c-statistics 0.76–0.78).

Repeating these analysis *without* adjustment for age and sex, c-statistics of the new indices were 0.05, 0.05, and 0.05 higher than the Charlson, Elixhauser, and FCI for predicting ADL decline; 0.05, 0.04, and 0.01 higher for predicting IADL decline; 0.04, 0.04, and 0.03 higher for predicting hospitalization; and 0.03, 0.02, and 0.09 higher for predicting death, respectively (Supplementary Table S15). Age and sex are part of the CMS-HCC model, thus precluding a similar comparison for this model.

In decision curve analysis (which adjusted for age and sex), different indices had similar net benefit across a range of decision thresholds for outcomes of IADL decline and death (Figure 3). For outcomes of ADL decline and hospitalization, the new indices yielded

a slightly higher net benefit than legacy indices at intermediate and/or higher decision thresholds, but not at lower decision thresholds.

The integrated discrimination index (IDI) was in most cases slightly higher (better) for the new indices compared to each of the legacy indices for predicting ADL decline and hospitalization (1%–2% difference, $p < 0.05$ for each except CMS-HCC for hospitalization outcome; Supplementary Table S16). In contrast, in nearly all comparisons the IDI was <1% different between new versus legacy indices for predicting IADL decline and death. Results were similar for the net reclassification index (Supplementary Table S16).

Other observations

C-statistics for both new and legacy indices varied substantially between the derivation cohort and the primary and secondary validation cohorts (Supplementary Table S17). Most notably, c-statistics for outcomes of ADL decline, IADL decline, and death were substantially lower in both validation cohorts for all indices, with each outcome having a different pattern of c-statistic variation across cohorts. In contrast, the c-statistics for hospitalization were higher in the validation cohorts for all indices, particularly the legacy ones. Additional information is shown in Supplementary Text and Table S18 and S19.

DISCUSSION

We developed and validated four new indices of multimorbidity using Medicare claims data, each tuned to a different outcome. In an independent validation sample, the new indices demonstrated good ability to discriminate between older adults who did versus did not develop ADL decline, hospitalization, and death over 2 years (c-statistics of 0.72, 0.72, and 0.77, respectively, each including adjustment for age and sex). Performance for IADL decline was more anemic (c-statistic 0.69). The new indices slightly outperformed legacy indices including the Charlson Comorbidity Index, Elixhauser Index, and Functional Comorbidity Index, with c-statistics of 0.01–0.03 higher across outcomes of interest after adjustment for age and sex, and performed similarly to the more complex CMS-HCC model. On Decision Curve Analysis, the new indices had slightly higher net benefit for outcomes of ADL decline and hospitalization at certain decision thresholds, but not others.

In this study, we sought to overcome the limitations of prior claims-based multimorbidity indices by incorporating features that we hypothesized could improve model performance, including consideration of geriatric syndromes, markers of disease severity, and disease-disease interactions.[24,26,28,32,63–70] Did we succeed? The answer is mixed. In an independent validation cohort—which allowed a fair comparison—after adjusting for age and sex our indices demonstrated c-statistics 0.01–0.03 higher for a variety of outcomes compared to indices such as the Charlson index that are commonly used in research. This is a statistically significant but only small to moderate improvement. When compared against the CMS-HCC model, a more complex approach that is used by payors for risk adjustment, c-statistics were nearly identical (difference of <0.01 for all outcomes). By means of context, after adjusting for age and sex the gain in c-statistic from including any index versus none ranged from 0.01 to 0.09 for each of the four outcomes (Figure 2). Similarly, on Decision

Curve Analysis that adjusted for age and sex, the new indices had small to no net benefit compared to legacy indices.

In subsidiary analyses that did not adjust for age and sex, the gap between new indices and legacy ones typically used for research widened, with c-statistics 0.03–0.05 higher for many (although not all) outcomes. This difference may be meaningful in settings that do not require age and sex adjustment.

Several strategies we hoped would yield gains in predictive performance did not bear fruit. Claims-based markers of disease severity and disease-disease interactions did not produce meaningful improvement in model performance. Further, c-statistics for both new and legacy indices varied substantially across our derivation cohort and two validation cohorts, particularly for (but not limited to) outcomes of ADL and IADL decline. The reasons for this may include differences in ADL and IADL outcome definitions between the derivation and validation cohorts, temporal changes in coding practices, and the peculiarities of local conditions. Overfitting in model development could have also contributed, although internal validation analyses and substantial swings in the c-statistics of legacy indices suggest this was far from the only cause. While the exact reasons remain elusive, the substantial variation in c-statistics across cohorts for all indices—a finding mirrored in prior scholarship[71,72]— suggests caution is warranted in interpreting the predictive power of any given index.

More generally, the limited magnitude of improvement over legacy approaches that we achieved in our indices, particularly after adjustment for age and sex, suggests there may be limited opportunity to wring further improvement out of claims-based, prognostically-oriented multimorbidity assessment for general populations of older adults. This interpretation is supported by prior scholarship, and in part reflects the crude nature of claims data.[3,73] It also likely reflects the intrinsic limitations of creating an index for a broad and heterogeneous population. Even when certain conditions such as metastatic cancer are highly predictive of poor outcomes, only a fraction of people will have one of these conditions and so their predictive contribution on a population level is diminished. Incorporating other markers that predict poor outcomes, for example social determinants of health or purchase of durable medical equipment (as has been used in claims-based frailty and function-related indicators), holds substantial promise to improve our ability to predict future health outcomes.[30,31,48–50,74] However, such elements go beyond the purpose of a multimorbidity index, which typically aims to capture a single aspect of health status—the presence of multiple chronic conditions—and then apply it in conjunction with contextual factors to fully characterize a person's risk of adverse outcomes.[3,75] The interaction of these factors is an interesting area for future study—for example, should multimorbidity indices be modified not only for different outcomes but for people with different levels of access to health care and socioeconomic characteristics—although such questions need to be approached carefully to avoid unintended harmful consequences for historically underserved populations.[75]

If one does choose to use our indices, it is important to remember that they were primarily tested in cohorts that excluded people who died. Thus, the best use of the ADL decline, IADL decline, and hospitalization indices is in conjunction with the index to predict death

—for example, to apply both the death and ADL decline index to a given person to first characterize his or her likelihood of death in the next 2 years, and if the person survives, his or her likelihood of experiencing functional decline. It is also important to recognize that our approach was designed to identify associations, not prove causality. Thus, inclusion of individual conditions and their weights within each index—and variations across the four indices—should not be interpreted as causal or imply that certain conditions are more important than others.

Our study has several limitations. Our indices were developed with ICD-9-CM codes on data which are no longer recent, and it is unclear how well they would perform if adapted to ICD-10 and health records today. However, other studies of the ICD-9 to -10 transition suggest that performance of models such as ours are typically preserved.[12,29,76] There is also little reason to suspect that the relationship between multimorbidity and outcomes has meaningfully changed over the last decade, although results from our secondary validation cohort do suggest that predictive performance of both new and legacy indices vary over time and the population being assessed. Markers of disease severity were intentionally broad to allow application to multiple conditions, and so we did not assess markers highly tailored to individual conditions such as have been used in disease-specific studies. Our methods for identifying conditions relied on Medicare fee-for-service data and thus have uncertain applicability for people enrolled in Medicare Advantage plans or for younger adults. It is also difficult to know how much the limitations of our indices and legacy ones reflect the imperfect predictive power of multimorbidity versus limitations of claims data; prior studies comparing the predictive power of chart-review and claims-based multimorbidity indices have yielded conflicting results, although notably some have found similar predictive power for these two approaches even though agreement between chart review and claims data was fair to poor.[77–83]

It is possible that certain data-driven approaches would have yielded different or better indices; however we chose not to pursue these methods as they are susceptible to overfitting and often lack the transparency, clinical face validity, and simplicity that we felt end-users would need to trust and use the results of our research. While our sample size was insufficient to distinguish very small differences in performance of the models we tested, it was sufficient to detect clinically meaningful differences. Finally, it is important to remember that the impact of multimorbidity on people and their well-being extends far beyond its association with future outcomes such as death or functional decline that were the focus of this study.

In summary, a new series of claims-based indices outperformed legacy ones, but generally only by a small amount, and with predictive performance for both new and legacy indices varying substantially when applied to different cohorts. Failure to achieve a major improvement despite use of innovative methods suggests that intrinsic limitations in claims-based, prognostically-oriented multimorbidity measurement may be difficult to overcome. Choice among the various indices now available and new scholarship to improve measurement of multimorbidity remain important but merit a generous dose of humility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

SPONSOR'S ROLE

Study sponsors did not have a role in the design, methods, data ascertainment, analysis, preparation, or approval for publication of this article.

REFERENCES

1. Salive ME, Suls J, Farhat T, Klabunde CN. National Institutes of Health advancing multimorbidity research. Med Care. 2021;59(7):622–624. [PubMed: 33900269]

2. Simard M, Rahme E, Calfat AC, Sirois C. Multimorbidity measures from health administrative data using ICD system codes: a systematic review. Pharmacoepidemiol Drug Saf. 2022;31(1):1–12. [PubMed: 34623723]

3. Stirland LE, González-Saavedra L, Mullin DS, Ritchie CW, Muniz-Terrera G, Russ TC. Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice. BMJ. 2020;368:m160. [PubMed: 32071114]

4. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373–383. [PubMed: 3558716]

5. de Groot V, Beckerman H, Lankhorst GJ, Bouter LM. How to measure comorbidity. A critical review of available methods. J Clin Epidemiol. 2003;56(3):221–229. [PubMed: 12725876]

6. Extermann M. Measurement and impact of comorbidity in older cancer patients. Crit Rev Oncol Hematol. 2000;35(3):181–200. [PubMed: 10960800]

7. Klabunde CN, Warren JL, Legler JM. Assessing comorbidity using claims data: an overview. Med Care. 2002;40(8 Suppl):IV-26–35.

8. Yurkovich M, Avina-Zubieta JA, Thomas J, Gorenchtein M, Lacaille D. A systematic review identifies valid comorbidity indices derived from administrative health data. J Clin Epidemiol. 2015;68(1):3–14. [PubMed: 25441702]

9. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care. 1998;36(1):8–27. [PubMed: 9431328]

10. Wei MY, Kabeto MU, Langa KM, Mukamal KJ. Multimorbidity and physical and cognitive function: performance of a new multimorbidity-weighted index. J Gerontol A Biol Sci Med Sci. 2018;73(2):225–232. [PubMed: 28605457]

11. Wei MY, Luster JE, Ratz D, Mukamal KJ, Langa KM. Development, validation, and performance of a new physical functioning-weighted multimorbidity index for use in administrative data. J Gen Intern Med. 2021;36:2427–2433. [PubMed: 33469748]

12. Sun JW, Rogers JR, Her Q, et al. Adaptation and validation of the combined comorbidity score for ICD-10-CM. Med Care. 2017;55(12):1046–1051. [PubMed: 29087983]

13. Suls J, Bayliss EA, Berry J, et al. Measuring multimorbidity: selecting the right instrument for the purpose and the data source. Med Care. 2021;59(8):743–756. [PubMed: 33974576]

14. Sharabiani MT, Aylin P, Bottle A. Systematic review of comorbidity indices for administrative data. Med Care. 2012;50(12):1109–1118. [PubMed: 22929993]

15. Centers for Medicare & Medicaid Services. Risk Adjustment. Accessed 17 July, 2022. https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors

16. Yancik R, Havlik RJ, Wesley MN, et al. Cancer and comorbidity in older patients: a descriptive profile. Ann Epidemiol. 1996;6(5):399–412. [PubMed: 8915471]

17. Ryan A, Wallace E, O'Hara P, Smith SM. Multimorbidity and functional decline in community-dwelling adults: a systematic review. Health Qual Life Outcomes. 2015;13(1):168. [PubMed: 26467295]

18. Marengoni A, von Strauss E, Rizzuto D, Winblad B, Fratiglioni L. The impact of chronic multimorbidity and disability on functional decline and survival in elderly persons. A community-based, longitudinal study. J Intern Med. 2009;265(2):288–295. [PubMed: 19192038]

19. Holman CD, Preen DB, Baynham NJ, Finn JC, Semmens JB. A multipurpose comorbidity scoring system performed better than the Charlson index. J Clin Epidemiol. 2005;58(10):1006–1014. [PubMed: 16168346]

20. Satariano WA, Ragland DR. The effect of comorbidity on 3-year survival of women with primary breast cancer. Ann Intern Med. 1994;120(2):104–110. [PubMed: 8256968]

21. Busija L, Osborne RH, Roberts C, Buchbinder R. Systematic review showed measures of individual burden of osteoarthritis poorly capture the patient experience. J Clin Epidemiol. 2013;66(8):826–837. [PubMed: 23810023]

22. Huntley AL, Johnson R, Purdy S, Valderas JM, Salisbury C. Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide. Ann Fam Med. 2012;10(2):134–141. [PubMed: 22412005]

23. Miller MD, Paradis CF, Houck PR, et al. Rating chronic medical illness burden in geropsychiatric practice and research: application of the cumulative illness rating scale. Psychiatry Res. 1992;41(3):237–248. [PubMed: 1594710]

24. Parkerson GR Jr, Broadhead WE, Tse CK. The Duke severity of illness checklist (DUSOI) for measurement of severity and comorbidity. J Clin Epidemiol. 1993;46(4):379–393. [PubMed: 8483003]

25. Rozzini R, Frisoni GB, Ferrucci L, et al. Geriatric index of comorbidity: validation and comparison with other measures of comorbidity. Age Ageing. 2002;31(4):277–285. [PubMed: 12147566]

26. Bayliss EA, Ellis JL, Steiner JF. Seniors' self-reported multimorbidity captured biopsychosocial factors not incorporated into two other data-based morbidity measures. J Clin Epidemiol. 2009;62(5):550–557 e551. [PubMed: 18757178]

27. van Seben R, Covinsky KE, Reichardt LA, et al. Insight into the posthospital syndrome: a 3-month longitudinal follow up on geriatric syndromes and their association with functional decline, readmission, and mortality. J Gerontol A Biol Sci Med Sci. 2020;75(7):1403–1410. [PubMed: 32072168]

28. Inouye SK, Studenski S, Tinetti ME, Kuchel GA. Geriatric syndromes: clinical, research, and policy implications of a core geriatric concept. J Am Geriatr Soc. 2007;55(5):780–791. doi:10.1111/j.1532-5415.2007.01156.x [PubMed: 17493201]

29. Quan H, Sundararajan V, Halfon P, et al. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130–1139. [PubMed: 16224307]

30. Kim DH, Schneeweiss S. Measuring frailty using claims data for pharmacoepidemiologic studies of mortality in older adults: evidence and recommendations. Pharmacoepidemiol Drug Saf. 2014;23(9):891–901. [PubMed: 24962929]

31. Faurot KR, Jonsson Funk M, Pate V, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. Pharmacoepidemiol Drug Saf. 2015;24(1):59–66. [PubMed: 25335470]

32. Rosen A, Wu J, Chang BH, Berlowitz D, Ash A, Moskowitz M. Does diagnostic information contribute to predicting functional decline in long-term care? Med Care. 2000;38(6):647–659. [PubMed: 10843312]

33. Healthcare Cost and Utilization Project Clinical Classification Software. Accessed 11 May, 2021. http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

34. Centers for Medicare & Medicaid Services. Chronic Conditions Data Warehouse. Accessed 11 May, 2021. https://www.ccwdata.org/web/guest/home

35. Magnan EM, Bolt DM, Greenlee RT, Fink J, Smith MA. Stratifying patients with diabetes into clinically relevant groups by combination of chronic conditions to identify gaps in quality of care. Health Serv Res. 2018;53(1):450–468. [PubMed: 27861829]

36. Boyd CM, Weiss CO, Halter J, Han KC, Ershler WB, Fried LP. Framework for evaluating disease severity measures in older adults with comorbidity. J Gerontol A Biol Sci Med Sci. 2007;62(3):286–295. [PubMed: 17389726]

37. Fleming ST, Sabatino SA, Kimmick G, et al. Developing a claim-based version of the ACE-27 comorbidity index: a comparison with medical record review. Med Care. 2011;49(8):752–760. [PubMed: 21490514]

38. Melfi C, Holleman E, Arthur D, Katz B. Selecting a patient characteristics index for the prediction of medical outcomes using administrative claims data. J Clin Epidemiol. 1995;48(7):917–926. [PubMed: 7782800]

39. Abudagga A, Sun SX, Tan H, Solem CT. Exacerbations among chronic bronchitis patients treated with maintenance medications from a US managed care population: an administrative claims data analysis. Int J Chron Obstruct Pulmon Dis. 2013;8:175–185. [PubMed: 23589684]

40. Rizzo A, Jing B, Shah S, Steinman M. Can markers of disease severity improve the predictive power of claims-based multimorbidity indices? [Conference abstract]. J Am Geriatr Soc. 2021;69(Suppl 1):S214.

41. Weller SC, Romney AK. Systematic Data Collection. Sage Publications; 1988.

42. Thomas SL, Heck RH. Analysis of large-scale secondary data in higher education research: potential perils associated with complex sampling designs. Res High Educ. 2001;42:517–540.

43. Hastie T, Tibshirani R, Tibshirani R. Best subset, forward stepwise or lasso? Analysis and recommendations based on extensive comparisons. Stat Sci. 2020;35(4):579–592.

44. Sullivan LM, Massaro JM, D'Agostino RB Sr. Presentation of multivariate data for clinical use: the Framingham study risk score functions. Stat Med. 2004;23(10):1631–1660. [PubMed: 15122742]

45. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. J Clin Epidemiol. 2001;54(8):774–781. [PubMed: 11470385]

46. van Walraven C, Austin PC, Jennings A, Quan H, Forster AJ. A modification of the Elixhauser comorbidity measures into a point system for hospital death using administrative data. Med Care. 2009;47(6):626–633. [PubMed: 19433995]

47. Kumar A, Graham JE, Resnik L, et al. Examining the association between comorbidity indexes and functional status in hospitalized Medicare fee-for-service beneficiaries. Phys Ther. 2016;96(2):232–240. [PubMed: 26564253]

48. Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring frailty in Medicare data: development and validation of a claims-based frailty index. J Gerontol A Biol Sci Med Sci. 2018;73(7):980–987. [PubMed: 29244057]

49. Davidoff AJ, Zuckerman IH, Pandya N, et al. A novel approach to improve health status measurement in observational claims-based studies of cancer treatment and outcomes. J Geriatr Oncol. 2013;4(2):157–165. [PubMed: 23795223]

50. Segal JB, Chang HY, Du Y, Walston JD, Carlson MC, Varadhan R. Development of a claims-based frailty indicator anchored to a well-established frailty phenotype. Med Care. 2017;55(7):716–722. [PubMed: 28437320]

51. Moons KG, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162(1):W1–W73. [PubMed: 25560730]

52. Localio AR, Goodman S. Beyond the usual prediction accuracy metrics: reporting results for clinical decision making. Ann Intern Med. 2012;157(4):294–295. [PubMed: 22910942]

53. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology. 2010;21(1):128–138. [PubMed: 20010215]

54. Barnes DE, Mehta KM, Boscardin WJ, et al. Prediction of recovery, dependence or death in elders who become disabled during hospitalization. J Gen Intern Med. 2013;28(2):261–268. [PubMed: 23054919]

55. Van Calster B, Van Belle V, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW. Extending the c-statistic to nominal polytomous outcomes: the polytomous discrimination index. Stat Med. 2012;31(23):2610–2626. [PubMed: 22733650]

56. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ. 2016;352:i6. [PubMed: 26810254]

57. Kotwal AA, Lee SJ, Dale W, Boscardin WJ, Waite LJ, Smith AK. Integration of an objective cognitive assessment into a prognostic index for 5-year mortality prediction. J Am Geriatr Soc. 2020;68(8):1796–1802. doi:10.1111/jgs.16451 [PubMed: 32356919]

58. Vickers AJ, Pepe M. Does the net reclassification improvement help us evaluate models and markers? Ann Intern Med. 2014; 160(2):136–137. [PubMed: 24592500]

59. Pepe MS. Problems with risk reclassification methods for evaluating prediction models. Am J Epidemiol. 2011;173(11):1327–1335. [PubMed: 21555714]

60. Pepe MS, Fan J, Feng Z, Gerds T, Hilden J. The net reclassification index (NRI): a misleading measure of prediction improvement even with independent test data sets. Stat Biosci. 2015;7(2):282–295. [PubMed: 26504496]

61. Austin PC, Harrell FE Jr, Steyerberg EW. Predictive performance of machine and statistical learning methods: impact of data-generating processes on external validity in the "large N, small p" setting. Stat Methods Med Res. 2021;30(6):1465–1483. [PubMed: 33848231]

62. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015;162(1):55–63. [PubMed: 25560714]

63. Chassin MR. Getting better at measuring hospital mortality. JAMA Intern Med. 2020;180(3):355–356. [PubMed: 31860012]

64. Silva GC, Jiang L, Gutman R, et al. Mortality trends for veterans hospitalized with heart failure and pneumonia using claims-based vs clinical risk-adjustment variables. JAMA Intern Med. 2020;180(3):347–355. [PubMed: 31860015]

65. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. J Am Geriatr Soc. 1968;16(5):622–626. doi:10.1111/j.1532-5415.1968.tb02103.x [PubMed: 5646906]

66. Kaplan MH, Feinstein AR. The importance of classifying initial co-morbidity in evaluating the outcome of diabetes mellitus. J Chronic Dis. 1974;27(7–8):387–404. [PubMed: 4436428]

67. Greenfield S, Apolone G, McNeil BJ, Cleary PD. The importance of co-existent disease in the occurrence of postoperative complications and one-year recovery in patients undergoing total hip replacement. Comorbidity and outcomes after hip replacement. Med Care. 1993;31(2):141–154. [PubMed: 8433577]

68. Dubois MF, Dubuc N, Kroger E, Girard R, Hebert R. Assessing comorbidity in older adults using prescription claims data. J Pharm Health Serv Res. 2010;1(4):157–165.

69. Groll DL, To T, Bombardier C, Wright JG. The development of a comorbidity index with physical function as the outcome. J Clin Epidemiol. 2005;58(6):595–602. [PubMed: 15878473]

70. Greenfield S, Sullivan L, Dukes KA, Silliman R, D'Agostino R, Kaplan SH. Development and testing of a new measure of case mix for use in office practice. Med Care. 1995;33(4 Suppl):AS47–AS55. [PubMed: 7723461]

71. Schneeweiss S, Maclure M. Use of comorbidity scores for control of confounding in studies using administrative databases. Int J Epidemiol. 2000;29(5):891–898. [PubMed: 11034974]

72. Marengoni A, Angleman S, Melis R, et al. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev. 2011;10(4):430–439. [PubMed: 21402176]

73. Bastian LA, Brandt CA, Justice AC. Measuring multimorbidity: a risky business. J Gen Intern Med. 2017;32(9):959–960. [PubMed: 28653231]

74. Chrischilles E, Schneider K, Wilwert J, et al. Beyond comorbidity: expanding the definition and measurement of complexity among older adults using administrative claims data. Med Care. 2014;52(Suppl 3):S75–S84. [PubMed: 24561763]

75. Bayliss EA, Bonds DE, Boyd CM, et al. Understanding the context of health for persons with multiple chronic conditions: moving from what is the matter to what matters. Ann Fam Med. 2014;12(3):260–269. [PubMed: 24821898]

76. Toson B, Harvey LA, Close JC. New ICD-10 version of the multipurpose Australian comorbidity scoring system outperformed Charlson and Elixhauser comorbidities in an older population. J Clin Epidemiol. 2016;79:62–69. [PubMed: 27101889]

77. De Giorgi A, Di Simone E, Cappadona R, et al. Validation and comparison of a modified Elixhauser index for predicting inhospital mortality in Italian internal medicine wards. Risk Manag Healthc Policy. 2020;13:443–451. [PubMed: 32547275]

78. Hua-Gen Li M, Hutchinson A, Tacey M, Duke G. Reliability of comorbidity scores derived from administrative data in the tertiary hospital intensive care setting: a cross-sectional study. BMJ Health Care Inform. 2019;26:e000016. doi:10.1136/bmjhci-2009-000016

79. Hwang J, Chow A, Lye DC, Wong CS. Administrative data is as good as medical chart review for comorbidity ascertainment in patients with infections in Singapore. Epidemiol Infect. 2016; 144(9):1999–2005. [PubMed: 26758244]

80. Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H. A comparison of the Charlson Comorbidity Index derived from medical record data and administrative billing data. J Clin Epidemiol. 1999;52(2):137–142. [PubMed: 10201654]

81. Luthi JC, Troillet N, Eisenring MC, et al. Administrative data outperformed single-day chart review for comorbidity measure. International J Qual Health Care. 2007;19(4):225–231.

82. Stavem K, Hoel H, Skjaker SA, Haagensen R. Charlson Comorbidity Index derived from chart review or administrative data: agreement and prediction of mortality in intensive care patients. Clin Epidemiol. 2017;9:311–320. [PubMed: 28652813]

83. Susser SR, McCusker J, Belzile E. Comorbidity information in older patients at an emergency visit: self-report vs. administrative data had poor agreement but similar predictive validity. J Clin Epidemiol. 2008;61(5):511–515. [PubMed: 18394546]

Key points

- New measures of multimorbidity using claims data had moderate ability to predict ADL decline, IADL decline, hospitalization, and death.

- Performance of these indices was similar to or slightly better than legacy indices such as the Charlson Comorbidity Index.

Why does this paper matter?

There may be limited opportunity to improve measurement of multimorbidity in claims data over existing methods.
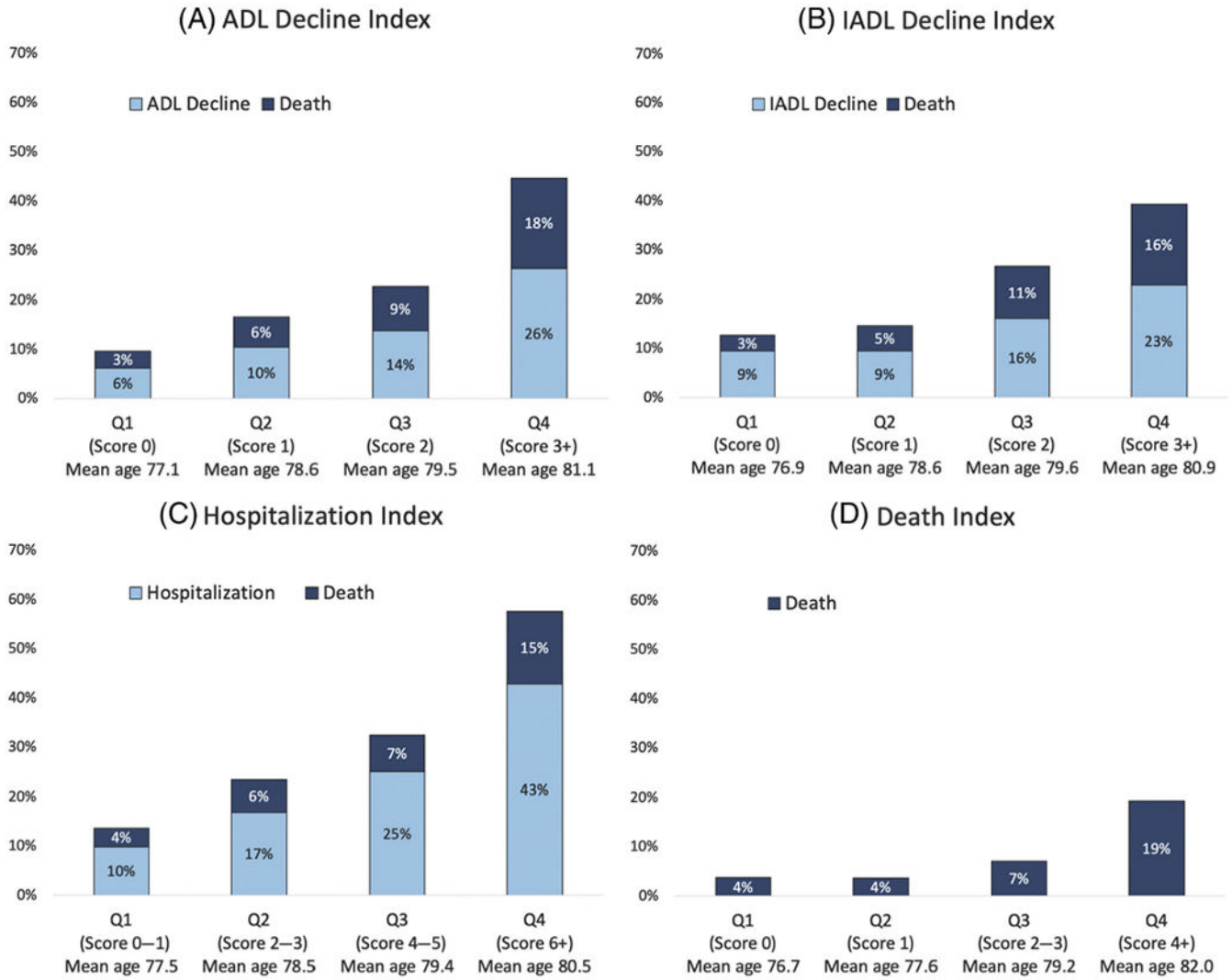
FIGURE 1.

Outcome rates in NHATS validation cohort per quartile of index score, four indices. Each panel shows the outcome rate per quartile of the respective index score (including adjustment for survey weights). Mean age of subjects in each quartile are shown in the horizontal axis labels. Panels A, B, and C each show two values: (1) the frequency of death (a competing risk for the other outcomes) per quartile of outcome score, and (2) among those who remained alive at 2 years and had follow-up data for the outcome of interest, the frequency of that outcome. This makes the denominator for the non-death outcome in Panels A, B, and C different than the denominator for the death outcome.
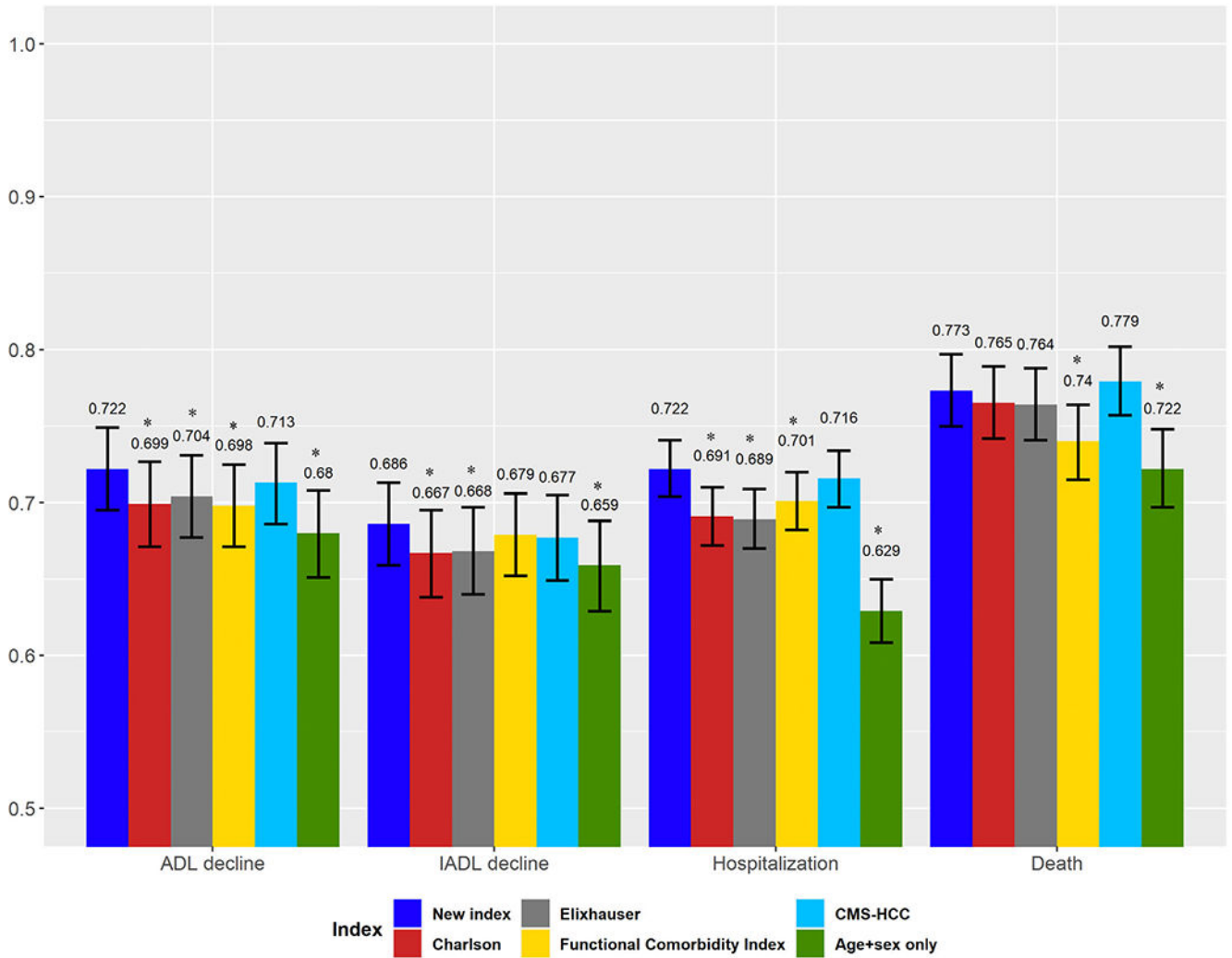
FIGURE 2.
C-statistics for novel indicates versus legacy indices, NHATS validation cohort. Figure shows c-statistics and associated 95% confidence intervals for models that include the index score, age, and sex. For outcomes with competing risk of death, c-statistics were calculated by excluding decedents from the cohort, thus eliminating death as a third outcome. Asterisk (*) indicates where c-statistic for a legacy index differs from c-statistic for the new index at $p < 0.05$ (see Supplementary Tables S14 and S17 for details).
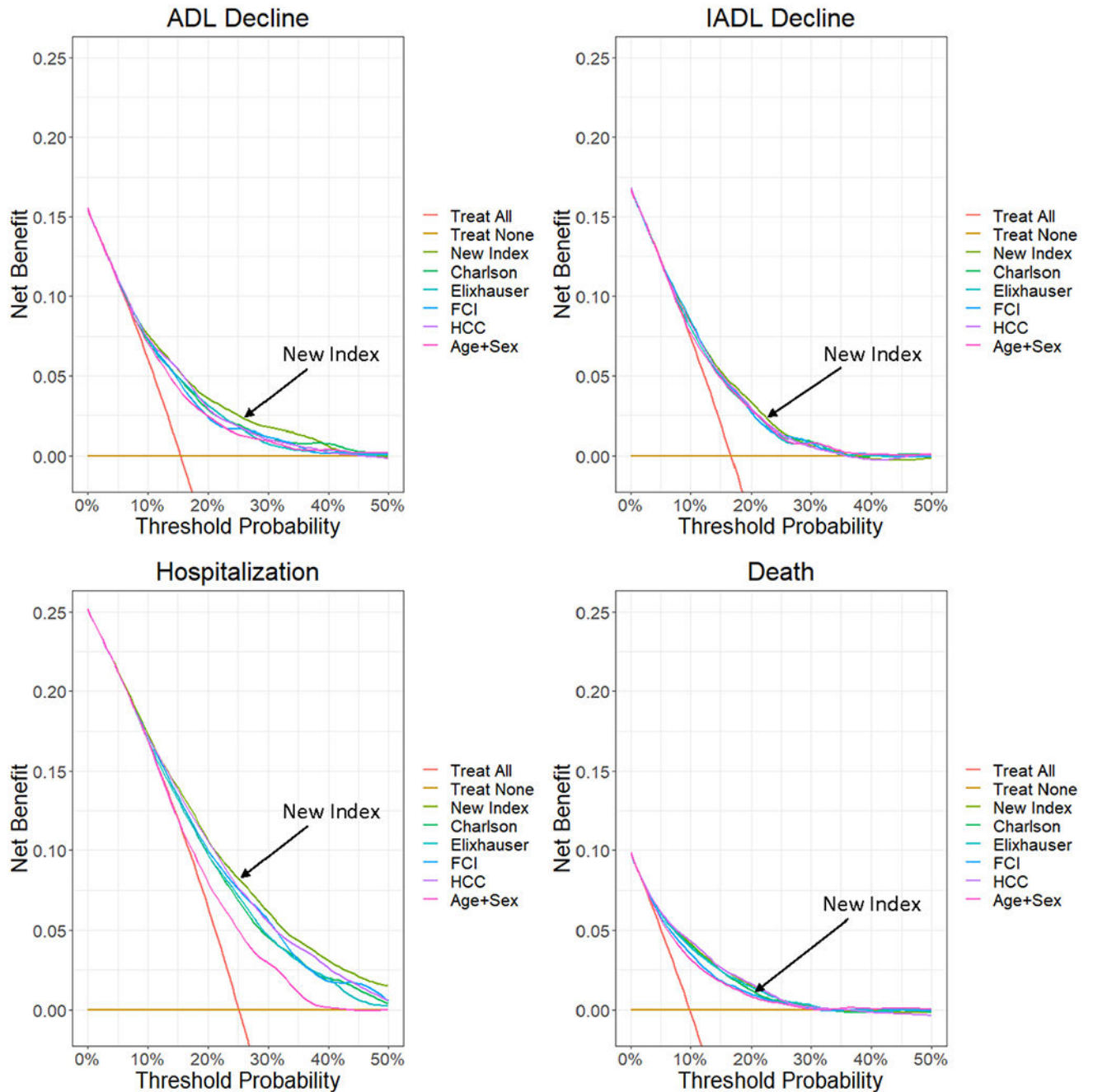
FIGURE 3.

Decision curve for new versus legacy indices. Figure shows decision curves for models containing age, sex, and new versus legacy indices applied to the NHATS validation cohort. (For the CMS-HCC, age and sex are already incorporated into the index). Separate panels show results for each of four outcomes. Within each panel, performance of each index at a given decision threshold (shown on the horizontal axis) is determined by the net benefit of the index at that threshold (shown on the vertical axis). For example, if a decision-maker wanted to refer older adults with 25% 2-year risk of ADL decline to a specialized program,

she would evaluate the top left panel to determine which multimorbidity index yields the highest net benefit at the 25% decision threshold for that outcome. Age + sex, model containing only age and sex (no multimorbidity index); FCI, functional comorbidity index; HCC, CMS-HCC model

TABLE 1

Characteristics of subjects in derivation and validation cohorts

	Derivation cohort (HRS) ($N = 5102$)	Primary validation cohort (NHATS) ($N = 4145$)
	Unweighted N (weighted %)	Unweighted N (weighted %)
Age (median, IQR)	76 (71–82)	75 (70–81)
Female	2962 (58.7%)	2427 (57.3%)
Race/ethnicity [a]		
White	4157 (86.9%)	3161 (88.5%)
African American	557 (6.6%)	812 (6.8%)
Hispanic/Latinx	302 (4.7%)	65 (1.4%)
Other	84 (1.8%)	107 (3.3%)
Lives alone	1478 (33.1%)	1456 (32.8%)
Dependency in any ADL [b]	518 (10.1%)	782 (15.4%)
Dependency in specific ADLs		
Bathing	338 (6.6%)	468 (8.7%)
Dressing	372 (7.0%)	538 (10.7%)
Transferring	173 (3.3%)	260 (4.9%)
Walking across room	245 (4.8%)	370 (7.0%)
Toileting	115 (2.2%)	173 (3.1%)
Eating	138 (2.8%)	236 (4.4%)
Dependency in any IADL	899 (17.9%)	1117 (21.5%)
Dependency in specific IADLs		
Preparing hot meal	377 (7.5%)	619 (11.5%)
Shopping for groceries	556 (10.9%)	943 (18.0%)
Managing money	403 (7.8%)	661 (12.1%)
Taking medications	195 (3.5%)	448 (7.9%)
Using telephone	291 (5.5%)	–

	Derivation cohort (HRS) (*N* = 5102)	Primary validation cohort (NHATS) (*N* = 4145)
	Unweighted *N* (weighted %)	Unweighted *N* (weighted %)
Cognitive impairment (based on Medicare claims)	553 (11.1%)	550 (10.9%)
Charlson Comorbidity Score (median IQR)	1 (0–3)	1 (0–3)
Hospitalized in previous year	976 (19.3%)	769 (17.1%)

Note: Subject characteristics are from the baseline visit (2010 for HRS, 2011 for NHATS). Results shown are for the death and hospitalization cohorts. Cohorts for ADL and IADL outcomes had fewer subjects but similar characteristics (Supplementary Table S07). All results except raw *N*'s are adjusted for sampling weights; because of these percents do not match dividing the raw numerator by the raw denominator.

[a] Race and ethnicity were assessed as separate variables in HRS and summarized as a single, mutually exclusive variable in NHATS, and are thus not directly comparable. Both inquired about identification with "Hispanic or Latino" ethnicity. For the HRS cohort, categories shown above are white/not Hispanic or Latinx, African American/not Hispanic or Latinx, other/not Hispanic or Latinx, and Hispanic. For the NHATs cohort, Hispanic or Latinx ethnicity is treated as mutually exclusive of other racial categories.

[b] ADLs and IADLs were assessed differently in the HRS and NHATS datasets (see Supplementary Text S02).

TABLE 2

Multimorbidity indices for each of four outcomes

ADL decline		IADL decline		Hospitalization		Death	
Condition	Points	Condition	Points	Condition	Points	Condition	Points
Metastatic cancer	3	Cognitive impairment	3	Tobacco use	2	Metastatic cancer	3
Parkinson's disease	3	Tobacco use	2	Diabetes with complications	2	Cognitive impairment	2
Cognitive impairment	2	Parkinson's disease	2	Back pain and related disorders	2	Chronic hematologic malignancy	2
Heart failure	1	Heart failure	1	Venous thrombo-embolism	1	Heart failure	1
COPD	1	COPD	1	Heart failure	1	COPD	1
Osteoarthritis & related disorders	1	Fluid and electrolyte disorders	1	COPD	1	Tobacco use	1
Delirium	1	Diabetes with complications	1	Fluid and electrolyte disorders	1	Fluid and electrolyte disorders	1
Chronic malaise or fatigue	1	Ischemic heart disease	1	Falls	1	Osteoarthritis & related disorders	1
Peripheral neuropathy	1	Hearing impairment	1	Ischemic heart disease	1	Falls	1
		Abnormal gait or difficulty walking	1	Iron deficiency anemia	1	Iron deficiency anemia	1
		Weight loss, malnutrition, adult failure to thrive; or debility	2	Fibrotic lung disease, lung disease due to external agents, or other lower respiratory tract disease	1	Other or unspecified anemia	1
				Atrial fibrillation or flutter	1	Chronic skin ulcers	1
				Arrythmia other than atrial fibrillation/flutter	1		
				Degenerative nervous system conditions other than Parkinson's disease	1		
				Depression	1		
Total possible number of points	14	Total possible number of points	14	Total possible number of points	19	Total possible number of points	16
C-statistic—derivation cohort	0.774	C-statistic—derivation cohort	0.755	C-statistic—derivation cohort	0.707	C-statistic—derivation cohort	0.803

ADL decline		IADL decline		Hospitalization		Death	
Condition	Points	Condition	Points	Condition	Points	Condition	Points
C-statistic—primary validation cohort	0.722	C-statistic—primary validation cohort	0.722	C-statistic—primary validation cohort	0.686	C-statistic—primary validation cohort	0.773

Note: Several conditions were present across multiple indices, as follows. Present in all four indices: heart failure, chronic obstructive pulmonary disease (COPD). Present in three indices: cognitive impairment, tobacco use, fluid and electrolyte disorders. Present in two indices: metastatic cancer, Parkinson's Disease, diabetes with complications, osteoarthritis, ischemic heart disease, falls, iron deficiency anemia. C statistics were calculated using models that included age (as a linear variable), sex, and index score (i.e., sum of points). Models that evaluated c statistics for the outcomes of ADL decline, IADL decline, and hospitalization were tested in cohorts that excluded patients with the competing risk of death, as c statistics cannot be calculated for models with more than two outcomes. In testing done in the derivation cohort, parameter estimates for the association between clinical conditions and these outcomes were similar across these logistic regression models and the multinomial models that accounted for the competing risk of death. Total number of possible points per model is not proportional to a model's predictive power. Instead, in each model points per condition were normalized by dividing the parameter estimate for the condition by the parameter estimate for 5-year increase in age, and then rounded to the nearest integer. For example, if a condition was equally predictive of two separate outcomes, but the parameter estimate for a 5-year increase in age was twice as large for the first outcome as for the second outcome, the point scores for that condition would be half as large for the first outcome as for the second one. ICD-9-CM codes for conditions are shown in Supplementary Table S04A.