Title

Can markers of disease severity improve the predictive power of claims-based multimorbidity indices?

Journal

Journal of the American Geriatrics Society, 71(3)

Authors

Rizzo, Anael
Jing, Bocheng
Boscardin, W
et al.

Publication Date

2023-03-01

DOI

10.1111/jgs.18150

Peer reviewed

HHS Public Access

Author manuscript

J Am Geriatr Soc. Author manuscript; available in PMC 2024 March 01.

Published in final edited form as:

J Am Geriatr Soc. 2023 March ; 71(3): 845–857. doi:10.1111/jgs.18150.

Can markers of disease severity improve the predictive power of claims-based multimorbidity indices?

Anael Rizzo, BS,
David Geffen School of Medicine, University of California, Los Angeles

Bocheng Jing, MS,
Division of Geriatrics, University of California San Francisco and San Francisco VA Medical Center

W. John Boscardin, PhD,
Division of Geriatrics, University of California San Francisco and San Francisco VA Medical Center

Department of Epidemiology and Biostatistics, University of California, San Francisco

Sachin J. Shah, MD, MPH,
Section of Hospital Medicine, Massachusetts General Hospital

Michael A. Steinman, MD
Division of Geriatrics, University of California San Francisco and San Francisco VA Medical Center

Abstract

Background: Claims-based measures of multimorbidity, which evaluate the presence of a defined list of diseases, are limited in their ability to predict future outcomes. We evaluated whether claims-based markers of disease severity could improve assessments of multimorbid burden.

Methods: We developed 7 dichotomous markers of disease severity which could be applied to a range of diseases using claims data. These markers were based on the number of disease-associated outpatient visits, emergency department visits, and hospitalizations made by an individual over a defined interval; whether an individual with a given disease had outpatient visits to a specialist who typically treats that disease; and ICD-9 codes which connote more vs. less advanced or symptomatic manifestations of a disease. Using Medicare claims linked with Health and Retirement Study data, we tested whether including these markers improved ability to predict

Corresponding author: Michael Steinman, MD, 4150 Clement St, SFVAMC Box 181G, San Francisco CA 94121, (t) 415.221.4810 x23677, (f) 415.750.6641, mike.steinman@ucsf.edu, (twitter) @mikesteinman.

Author Contributions

Study concept and design: Rizzo, Jing, Boscardin, Shah, Steinman
Acquisition of subjects and/or data: Jing, Boscardin, Steinman
Analysis and interpretation of data: Rizzo, Jing, Boscardin, Shah, Steinman
Preparation of manuscript: Rizzo, Jing, Boscardin, Shah, Steinman

Disclosures

Dr. Steinman receives royalties from UpToDate and honoraria from the American Geriatrics Society for his service on the AGS Beers Criteria Update Expert Panel. Drs. Shah and Boscardin, Mr. Jing, and Ms. Rizzo have no disclosures.

ADL decline, IADL decline, hospitalization, and death compared to equivalent models which only included the presence or absence of diseases.

Results: Of 5,012 subjects, median age was 76 years and 58% were female. For a majority of diseases tested individually, adding each of the 7 severity markers yielded minimal increase in c-statistic (≤ 0.002) for outcomes of ADL decline and mortality compared to models considering only the presence vs absence of disease. Gains in predictive power were more substantial for a small number of individual diseases. Inclusion of the most promising marker in multi-disease multimorbidity indices yielded minimal gains in c-statistics (<0.001 – 0.007) for predicting ADL decline, IADL decline, hospitalization, and death compared to indices without these markers.

Conclusions: Claims-based markers of disease severity did not contribute meaningfully to the ability of multimorbidity indices to predict ADL decline, mortality, and other important outcomes.

Keywords

Illness severity; Multimorbid burden; Medicare; Health and Retirement Study; Aging

Introduction

Multimorbidity, the presence of multiple coexisting chronic health conditions, affects two-thirds of older Americans and increases with age.[1–3] It has substantial impacts on treatment complexity and outcomes including mortality, hospitalization, and functional status, and as such has important implications for clinical practice, research, and health policy.[4–7] Identifying efficient and effective methods to measure multimorbidity are thus critical to facilitate its use in each of these settings. Healthcare claims data provide a valuable tool for such efforts, and a number of indices for multimorbidity measurement have been developed or adapted for use with such data.[8–12]

Claims-based measures including the Charlson Comorbidity Index (CCI) and Elixhauser approaches typically only evaluate the presence or absence of a defined list of diseases.[8,13,14] However, we know clinically that not only the presence of a disease, but its severity can have major impacts on clinical outcomes.[15–17] Markers of disease severity – representing the extent of physiologic derangement, clinical impact, and resultant prognostic import of a given disease process - thus offer a promising way to improve multimorbidity indices by more accurately reflecting the total extent of illness burden and its impact on clinical outcomes. Disease severity markers have been incorporated into several non-claims-based methods evaluating multimorbid burden, primarily among studies using medical records, patient interviews, survey data, and chart review.[17,18] They have also been selectively applied to claims data for specific diseases and scenarios, such as hospitalization for acute stroke.[19–24] Yet such severity markers have not been applied to measures of overall multimorbid burden or in other broad-based systems. This is understandable, as in most cases diagnosis codes available in claims data provide only information on the presence or absence of disease and not its severity, and the types of claims-based markers that can be used for specific diseases (such as receipt of a specific procedure or medication) are not applicable across a range of diseases.

Our goal was to use Medicare claims data to construct markers of disease severity that could be applied across a range of conditions and would thus potentially be useful to improving assessment of multimorbid burden. For this purpose, we conceptualized disease severity as the degree of symptoms and signs of a given disease, with higher disease severity expected to confer worse prognosis for future clinical outcomes including functional decline and/or death. We hypothesized that we could identify markers of disease severity that were associated with worse prognosis, particularly when combined into a multimorbidity index that reflects the cumulative impact of diseases on these outcomes. We thus assessed whether inclusion of such severity markers could meaningfully improve prediction of patient outcomes when applied to individual conditions and to a summative multimorbidity index, compared with traditional methods of claims-based assessment that rely solely on measuring the presence vs absence of diseases.

Methods

This study used Medicare claims data linked to older adults in the Health and Retirement Study (HRS).[25] Approximately 90% of HRS subjects with Medicare consent to linkage of their Medicare data.[26] This study included adults in the HRS panel who were age 67 years or older in 2010, enrolled in Medicare Parts A and B at that time and over the 2 years prior, had consented to linkage of their Medicare data, and were community dwelling at the baseline (2010) HRS interview.

We evaluated 62 diseases and clinical syndromes (hereafter collectively termed “conditions”) which were identified using ICD-9 codes from outpatient and inpatient encounters using the Medicare Carrier, Outpatient, and Inpatient files. These conditions were derived from a starting list of 129 conditions obtained from existing indices of comorbidity, multimorbidity, and frailty, systematic reviews on the topic, additional literature, and expert input (see Supplementary Text 1). This list was reduced to 62 conditions using criteria from a larger study that aimed to create better methods of assessing multimorbid burden using claims data.[27] This process excluded conditions which were present in <1% of the study population, had strong conceptual overlap, or were not associated with outcomes of ADL decline or hospitalization at P<.20. The goal of creating this list was not to identify conditions that may be most amenable to or impacted by claims-based severity assessment. Rather, our goal was to identify conditions whose presence may contribute to adverse clinical outcomes (including functional decline, hospitalization, and/or death), and then to determine if adding claims-based severity markers to these conditions would yield greater ability to predict future clinical outcomes compared to simply considering the presence or absence of these conditions.

We considered a condition to be present if there were 1 or more corresponding ICD-9 codes over the 2 years prior to the 2010 HRS interview date. ICD-9 codes for each condition were based on existing coding schemas including the HCUP Clinical Classification System,[28] the Medicare Chronic Conditions Warehouse,[29] claims-based adaptations of Charlson and Elixhauser scores,[30] research by Kim,[31] Faurot,[32] and Rosen[33] (particularly for geriatric conditions), and where necessary, additional literature.

We conceptualized disease severity as the degree of symptoms and signs associated with a specific condition, which typically correlates with worse future outcomes. Prior research on claims-based markers of disease severity has been almost exclusively focused on single conditions using markers that are typically highly specific to that condition – for example, use of home oxygen and systemic corticosteroids to predict outcomes in people with chronic obstructive pulmonary disease.[21] Such disease-specific markers are not applicable across a range of conditions.

Our research team thus used group discussion, informed by extant literature and our research and clinical experience, to create a suite of 7 candidate markers that could potentially be used to assess severity of a variety of diseases in claims data. The first 4 included (1) emergency department visit for that condition in the past year, (2) hospitalization for the condition in the past year, (3) ≥ 4 outpatient visits on separate days for that condition in the past year, and (4) ≥ 1 visit for that condition in the past 6 months. We defined a visit being "for" that condition if the corresponding ICD-9 code was the primary reason for visit or the primary discharge diagnosis. The 5th marker was having had an outpatient visit with a provider who specializes in that condition over the past 6 months. To do this, we mapped each of the 62 conditions to the corresponding specialty using clinical judgement; for example, we designated cardiology as the specialty for heart failure (see Supplementary Tables 3 and 4). Conditions such as chronic malaise or fatigue could not be reliably mapped to a specialty. The 6th marker used ICD-9 codes which directly designated disease severity or complications, for example codes corresponding to "diabetes with renal manifestations" (ICD9: 250.4). Such distinctions were only available for 4 conditions (see Supplementary Table 5). The 7th marker consisted of a positive "hit" for severity on any of the previous 6 markers (i.e., "any of the above").

With limited exceptions these markers do not directly measure disease severity, but rather represent factors that are plausibly associated with severity. For example, time-based markers connote recency of a diagnosis being the focus of a clinical encounter. While this does not guarantee that the condition is severe, it is more likely to be clinically active and thus affecting the person's health than a condition which was coded in the more distant past but has not been the focus of a recent encounter.

For each subject, a condition was defined as absent (if there were no encounters with that condition in the previous 2 years) or present. If the condition was present, for each marker it was defined as severe if the marker definition was met and mild if the marker definition was not met. Thus, a given condition for a subject might be considered severe according to some markers and mild according to others.

The usefulness of these disease severity markers was assessed by their ability to enhance prediction: namely, whether models that contained these markers were better able to predict future clinical outcomes compared with equivalent models that only contained an indicator of whether a disease was present or absent. The outcomes of interest were functional decline, death, and acute care hospitalization, each assessed over 2 years following the 2010 baseline HRS interview. We defined decline in activities of daily living (ADLs) and instrumental activities of daily living (IADLs) as requiring help in a greater number of

ADLs or IADLs at 2 year follow-up compared to baseline. ADLs included bathing, dressing, transferring, walking across a room, toileting, and eating, and IADLs included preparing a hot meal, shopping for groceries, managing money, taking medications, and using the telephone; each were assessed using self or proxy report. Because functional status could not be assessed in people who did not complete the follow-up assessment, subjects who remained alive but without complete ADL data at baseline and/or follow-up were excluded from analyses where ADL decline was the outcome. Analogous methods were used for analyses involving the outcome of IADL decline. Death was assessed using National Death Index and Medicare Beneficiary Summary file, and part of the HRS RAND file. Acute care hospitalization was assessed using Medicare files; we excluded hospitalizations for psychiatric illness and elective procedures.[34]

While our ultimate goal was to evaluate performance of disease severity markers in multi-disease indices of multimorbidity, we started by testing each condition separately in limited multivariable models to evaluate how they performed for individual conditions and to guide assessment of which marker type was best suited to advance to multivariable testing. In these initial analyses, for a given condition we constructed a model that contained only age, sex, and an indicator variable denoting whether the condition was present or absent. We next created a companion model that substituted the “disease yes/no” indicator variable with a pair of indicator variables that indicated the presence of mild or severe disease (specified as “mild disease yes/no” and “severe disease yes/no”). We then compared c-statistics for these two models. We repeated this process for all 62 conditions, for each of the 7 disease severity markers, and for each of 2 outcomes (death and ADL decline). Because ADL decline is impacted by the competing risk of death, we constructed multinomial logistic regression models that accounted for this competing risk. The c-statistics for ADL decline were calculated from the related binary logistic regression models that excluded decedents and where the outcomes were ADL declined vs. not declined.[27] We focused these analyses on outcomes of ADL decline and death because they were of greatest a priori interest for assessing how disease severity markers could yield better methods of assessing multimorbidity burden using claims data.

We next sought to determine which of the 7 definitions performed best overall for both outcomes. We defined this as a synthesis of (1) number of conditions with at least marginally positive c-statistic gains in bivariate analysis, (2) the strength of those associations, (3) the complexity required to implement the definition in claims data, and (4) face validity. Inspection of graphs and summary statistics for our 62 conditions under the 7 definitions and 2 outcomes yielded no clear winner. Given this ambiguity, the research team discussed the decision of which definition was the “best” one to advance for further testing with an advisory panel convened for this study; Based on the research team’s judgement with support of the advisory panel we selected the definition of “>=4 outpatient visits for that condition in the past year” as the one we would advance for further testing.

Next, we tested the effect of incorporating this disease severity marker into a summative multimorbidity model that captured the effect of multiple diseases.[27] We had previously created 4 of these base multimorbidity models and summative multimorbidity indices, each corresponding to a separate outcome of interest. As reported elsewhere these predict several

clinical outcomes slightly to moderately better than legacy multimorbidity indices such as the Charlson Comorbidity Index.[27] Testing of disease severity markers was done on a pre-final version of these models that included conditions with negative parameter estimates (i.e.., which were protective against the outcome) and prior to conversion of parameters estimates into integers for point-based scoring. As described elsewhere, c-statistics for these pre-final models were very similar to c-statistics for the final version of the multimorbidity indices that excluded protective conditions and used a point-based scoring system.[27]

To evaluate whether disease severity markers could improve predictive power of our base models, we explored several approaches. First, we replaced each disease yes/no indicator in the base multimorbidity model with the corresponding disease-mild and disease-severe indicators for that condition. Second, we re-derived each model whereby each condition was entered either as a disease yes/no indicator and through a pair of disease severity indicators, based on results of the testing each disease individually. Third, we took the individual conditions for which disease severity markers yielded the highest gain in c-statistics, and appended disease severity markers for those conditions to our base models. Finally, for each subject we summed the number of conditions (out of the 62 candidate conditions) which were classified as severe and added this single variable to our base models. In all cases, we used the methods described above to compare the difference in c-statistics between the base models (which included only disease yes/no indicators) and the revised models (which also included disease severity indicators). The Spiegel-Halter Z score was used to assess model calibration.

Because results from the multivariable models did not demonstrate any meaningful gain in discrimination by incorporating markers of disease severity, we did not take the additional steps of externally validating the results, as further reduction in c-statistics would not change the conclusions of our work.

All analyses incorporate the respondents' baseline HRS survey weights. Analyses were performed using SAS version 9.4. This research was approved by the institutional review boards of the University of California, San Francisco and the San Francisco VA Medical Center.

Results

Our sample included 5,102 subjects with a mean age of 76 years; 58.7% were female (Table 1). Among this cohort, 25.5% were hospitalized and 8.9% died over the next 2 years. Among those who remained alive, 8.7% experienced ADL decline and 10.9% experienced IADL decline (after excluding 187 and 288 subjects who did not complete the follow-up survey, respectively).

The 62 diseases we assessed are shown in Table 2. The proportion of people with a given disease who were classified as severe varied widely between diseases and definitions of severity (see Supplementary Table 2). Across all 62 conditions, the median (interquartile range, IQR) of subjects whose disease was considered severe ranged from 0.6% (IQR 0.1% - 2.5%) when severe disease was defined by hospitalization for the condition in the past 12

months, to 24.9% (IQR 17.9% - 33.3%) when severe disease was defined as a positive hit on any of the other 6 severity markers. The median (IQR) frequency of severe disease across the 62 conditions was 5.5% (IQR, 2.1% - 13.3%) for severity marker 3, which defines severe disease as 4 or more outpatient visits for that condition in the past 12 months.

Figures 1 and 2 and Supplementary Figures 6A and 6B show the difference in c-statistics between the single-disease models that include only indicators for the presence vs absence of that disease vs. those that include indicators of the severity of that disease (all models also included age and sex). Overall, inclusion of disease severity markers showed a stronger effect for predicting the death outcome compared to the ADL outcome. Disease types with the highest gain in c-statistic from inclusion of disease severity markers included respiratory disorders other than asthma/COPD (c-statistic gain of 0.006 and 0.014 for ADL decline and death, respectively), lung diseases due to external agents, fibrotic lung disease, and related conditions (gain of 0.005 for ADL decline and 0.012 for death, respectively), non-metastatic cancer (gain of 0.021 for death), and diabetes (gain of 0.008 for ADL decline). For each of our 7 definitions of disease severity, the mean gain in c-statistic across the 62 conditions tested was <=0.001for death and for ADL decline (Supplementary Table 7).

No disease severity marker performed unambiguously better than the others. Based on the totality of results we selected severity marker #3 (four or more outpatient visits for a given condition in the past 12 months) as the most promising marker for further testing.

Next, we developed multivariable models using this disease severity marker and compared their performance to that of base multivariable models developed only with indicators of whether a disease was present or absent, plus age and sex. (These base models, each of which contained age, sex, and indicators for a variety of chronic conditions, correspond to a measure of multimorbid burden with separate models developed for outcomes of ADL decline, IADL decline, hospitalization, and death, as described in a separate publication).[27] When compared to those original multivariable models, the models that accounted for disease severity showed minimal changes in predictive power: a change in c-statistic of 0.003 for ADL decline, 0.002 for IADL decline, <0.001 for hospitalization, and 0.002 for death (Table 3). An alternate approach, whereby we appended disease severity markers to our original multivariable models, also yielded minimal gains in c statistics: 0.004 for ADL decline, 0.007 for IADL decline, 0.002 for hospitalization, and 0.004 for death. Calibration of all models was adequate (Spiegel-Halter Z score P value >0.05). Other variations of these approaches had similar results (Table 3).

Supplementary Table 8 provides additional information comparing the odds ratios for mild vs. severe forms of specific conditions. Within each outcome model (ADL decline, IADL decline, hospitalization, death), several conditions showed a statistically significant (P<.05) difference between odds ratios for mild and severe forms of that condition.

Discussion

Disease severity has important prognostic implications in clinical practice. Our study sought to identify candidate markers of disease severity that could be broadly applied to

claims data, and thus could be used to improve approaches to measuring multimorbidity using claims data for multiple different conditions. Gains in predictive power by adding these markers were appreciable for a small number of specific diseases when considered individually, particularly when predicting death. However, incorporating such markers in multimorbidity indices and the models from which they arose yielded no appreciable gain in their ability to predict ADL decline, mortality and other important outcomes.

While literature evaluating the impact of claims-based severity markers across multiple illnesses is remarkably limited, disease-specific studies have been conducted more frequently—often focusing on acute illness events—with results demonstrating improvements in predictive model performance. This comports with most approaches to the measurement of multimorbidity, whereby the validity of each of these measures is assessed by its ability to predict future outcomes. In a study by Ford et al., researchers developed a claims-based model to investigate the risk of in-hospital mortality for patients with severe sepsis.[35] This model incorporated measures of acute illness severity (mechanical ventilation, vasopressor medication, hemodialysis, and treatment in the ICU), and attained a C-statistic of 0.80 and 0.84. Similar studies have been conducted by Lagu and Schwarzkopf, with comparable predictive ability for their sepsis mortality models.[36,37] Simpson et al. developed a Stroke Administrative Severity Index (SASI) to predict 30-day post stroke outcomes including mortality and discharge to hospice.[38] Adding the SASI score to a model with age, sex, race and Charlson score improved the C-statistic from 0.72 to 0.77. Annavarapu and colleagues validated a model with good predictive ability that evaluated risk of severe chronic obstructive pulmonary disease (COPD) exacerbation, identifying history of severe exacerbation, Deyo-Charlson comorbidity score, COPD-related inpatient stays, and use of oxygen therapy as the strongest predictors associated with increased risk of exacerbation.[21]

Our attempts to measure severity using broad-based approaches that could apply to a wide array of diseases presented a different set of challenges. For several conditions, there were clinically meaningful and statistically significant differences in the association between mild vs. severe forms of disease with our outcomes of interest. However, given that these approaches were not tailored to the specific circumstances of single condition, they were less likely to yield major gains in predictive power for any given disease. Applying them to multivariable and index-based models of multimorbidity also presented opportunities and drawbacks. On one hand, small to moderate gains in predictive power across a series of diseases have potential to accrue to meaningful gains when many of these diseases are combined into single model. Conversely, gains in predictive power from certain diseases (as found in our single-disease analyses) can be diluted by the negligible effects of others. Moreover, each new piece of information (i.e., adding a new condition) in a multimorbidity model typically yields diminishing returns for predictive power – such that there may have been little opportunity to gain predictive power by adding disease severity to a model that already had multiple disease predictors. Finally, perhaps the markers we evaluated were too crude. While more refined measures are worth exploring, it may be that searching for a one-size-fits-all marker of disease severity based on claims data, especially outside the acute care setting, will not be fruitful.

There are some limitations to our study. Markers of disease severity were contingent on patients receiving care in healthcare settings, including a mix of inpatient and outpatient visits. Our measures thus to some extent entangle healthcare utilization with disease diagnosis. For instance, if a person was unable to attend outpatient visits due to social disadvantage and then became extremely ill and was hospitalized, this would be captured during hospitalization but not during the outpatient visit. This person would thus not be counted as having severe disease (or any disease at all) for markers that depend on diagnoses from outpatient claims. In addition, our models were developed using ICD-9 codes to identify diseases, and for marker #6 to identify severity. While ICD-10 codes offer different opportunities, the increased granularity of these codes typically reflects finer gradations between disease types and body parts affected, and not gradations of disease severity per se. Certain markers of disease severity that we tested were present in only a small proportion of people with a given condition, thus limiting the precision of estimates. We are unable to distinguish to what extent our markers measured disease severity vs. disease activity, although we suspect that these features are strongly (although not inextricably) linked for many common clinical conditions. Finally, we validated our measures of disease severity by their ability to predict future outcomes including functional decline and death, and we are unable to comment on their ability to predict current symptoms, health care costs, or other potential sequelae of heightened disease severity.

Our results showed that claims-based markers of disease severity did not improve the ability of multimorbidity indices to predict ADL decline, mortality and hospitalizations. These findings do not indicate that disease severity is not important – any clinician knows otherwise. Thus, the task at hand is to develop better ways of accounting for disease severity using claims data and to apply them wisely, recognizing that disease-specific approaches may be more likely to bear fruit than broad-based, multi-disease approaches.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank the members of the scientific advisory board including Drs. Mary Tinetti, Cynthia Boyd, Eve Kerr, Lillian Min, Denis Cortese, Mindy Fain, and Jodi Segal.

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.

Funding:

This research was funded by the National Institute of Aging (grants R01AG052041, K24AG049057, P01AG066605, and P30AG044281).

References:

1. Anderson G. Robert Wood Johnson Foundation. Chronic Care: Making the Case for Ongoing Care. 2010. https://www.rwjf.org/en/library/research/2010/01/chronic-care.html. Updated 1 January 2010. Accessed 6 June, 2021.

2. Centers for Medicare and Medicaid Services. Chronic conditions among Medicare beneficiaries, Chartbook, 2012 Edition. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Downloads/2012Chartbook.pdf. Published 2012. Accessed 22 August, 2021.

3. Boyd CM, Ritchie CS, Tipton EF, Studenski SA, Wieland D. From Bedside to Bench: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Comorbidity and Multiple Morbidity in Older Adults. Aging Clin Exp Res. 2008;20(3):181–188. [PubMed: 18594183]

4. Gijsen R, Hoeymans N, Schellevis FG, Ruwaard D, Satariano WA, van den Bos GA. Causes and consequences of comorbidity: a review. J Clin Epidemiol. 2001;54(7):661–674. [PubMed: 11438406]

5. Boyd CM, Fortin M. Future of multimorbidity research: how should understanding of multimorbidity inform health system design? Public Health Rev. 2010;32(2):451–474.

6. Lee TA, Shields AE, Vogeli C, Gibson TB, Woong-Sohn M, Marder WD, Blumenthal D, Weiss KB. Mortality rate in veterans with multiple chronic conditions. J Gen Intern Med. 2007;22 Suppl 3:403–407. [PubMed: 18026809]

7. Ritchie CS, Zulman DM. Research priorities in geriatric palliative care: multimorbidity. J Palliat Med. 2013;16(8):843–847. [PubMed: 23777331]

8. Austin SR, Wong Y-N, Uzzo RG, Beck JR, Egleston BL. Why Summary Comorbidity Measures Such As the Charlson Comorbidity Index and Elixhauser Score Work. Med Care. 2015;53(9):e65–72. [PubMed: 23703645]

9. 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]

10. Suls J, Bayliss EA, Berry J, Bierman AS, Chrischilles EA, Farhat T, Fortin M, Koroukian SM, Quinones A, Silber JH, Ward BW, Wei M, Young-Hyman D, Klabunde CN. Measuring Multimorbidity: Selecting the Right Instrument for the Purpose and the Data Source. Med Care. 2021;59(8):743–756. [PubMed: 33974576]

11. 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]

12. 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.

13. Li B, Evans D, Faris P, Dean S, Quan H. Risk adjustment performance of Charlson and Elixhauser comorbidities in ICD-9 and ICD-10 administrative databases. BMC Health Serv Res. 2008;8:12. [PubMed: 18194561]

14. Kumar A, Graham JE, Resnik L, Karmarkar AM, Deutsch A, Tan A, Al Snih S, Ottenbacher KJ. 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]

15. Boyd CM, Weiss CO, Halter J, Han KC, Ershler WB, Fried LP. Framework for evaluating disease severity measures in older adults with comorbidity. The Journals of Gerontology Series A, Biological Sciences and Medical Sciences. 2007;62(3):286–295. [PubMed: 17389726]

16. Crabtree HL, Gray CS, Hildreth AJ, O'Connell JE, Brown J. The Comorbidity Symptom Scale: a combined disease inventory and assessment of symptom severity. J Am Geriatr Soc. 2000;48(12):1674–1678. [PubMed: 11129761]

17. Fleming ST, Sabatino SA, Kimmick G, Cress R, Wu X-C, Trentham-Dietz A, Huang B, Hwang W, Liff J. 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]

18. Marengoni A, Angleman S, Melis R, Mangialasche F, Karp A, Garmen A, Meinow B, Fratiglioni L. Aging with multimorbidity: a systematic review of the literature. Ageing Res Rev. 2011;10(4):430–439. [PubMed: 21402176]

19. Sung S-F, Hsieh C-Y, Kao Yang Y-H, Lin H-J, Chen C-H, Chen Y-W, Hu Y-H. Developing a stroke severity index based on administrative data was feasible using data mining techniques. J Clin Epidemiol. 2015;68(11):1292–1300. [PubMed: 25700940]

20. Hung LC, Sung SF, Hsieh CY, Hu YH, Lin HJ, Chen YW, Yang YK, Lin SJ. Validation of a novel claims-based stroke severity index in patients with intracerebral hemorrhage. J Epidemiol. 2017;27(1):24–29. [PubMed: 28135194]
21. Annavarapu S, Goldfarb S, Gelb M, Moretz C, Renda A, Kaila S. Development and validation of a predictive model to identify patients at risk of severe COPD exacerbations using administrative claims data. Int J Chron Obstruct Pulmon Dis. 2018;13:2121–2130. [PubMed: 30022818]
22. Stanford RH, Nag A, Mapel DW, Lee TA, Rosiello R, Schatz M, Vekeman F, Gauthier-Loiselle M, Merrigan JFP, Duh MS. Claims-based risk model for first severe COPD exacerbation. The American journal of managed care. 2018;24(2):e45–e53. [PubMed: 29461849]
23. Yu T-H, Fu P-K, Tung Y-C. Using medication utilization information to develop an asthma severity classification model. BMC Med Inform Decis Mak. 2017;17(1):177. [PubMed: 29262809]
24. Jacob C, Haas JS, Bechtel B, Kardos P, Braun S. Assessing asthma severity based on claims data: a systematic review. The European Journal of Health Economics. 2017;18(2):227–241. [PubMed: 26931557]
25. HRS Sample Evolution: 1992–1998. Survey Research Center, Institute for Social Research, University of Michigan. https://hrs.isr.umich.edu/publications/biblio/9047. Published 2008. Accessed 21 August, 2021.
26. Weir D, Langa K. How Well Can Medicare Records Identify Seniors with Cognitive Impairment Needing Assistance with Financial Management? University of Michigan Retirement Research Center (MRRC) Working Paper, WP 2018–391 In. Ann Arbor, MI2018.
27. Steinman MA, Jing B, Shah SJ, Rizzo A, Lee SJ, Covinsky KE, Ritchie CS, Boscardin WJ. Development and validation of novel multimorbidity indices for older adults. J Amer Geriatr Soc (In Print). 2022.
28. Agency for Healthcare Research and Quality. User Guide: Clinical Classifications Software (CCS) for ICD-9-CM. https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp. Published 2015. Accessed.
29. Center for Medicare and Medicaid Services. Chronic Conditions Data Warehouse. https://www2.ccwdata.org/web/guest/data-dictionaries. Accessed 21 August, 2021.
30. Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi J-C, Saunders LD, Beck CA, Feasby TE, Ghali WA. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med Care. 2005;43(11):1130–1139. [PubMed: 16224307]
31. 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]
32. Faurot KR, Jonsson Funk M, Pate V, Brookhart MA, Patrick A, Hanson LC, Castillo WC, Stürmer T. 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]
33. 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]
34. Yale New Haven Health Services Corporation/Center for Outcomes Research & Evaluation (YNHHSC/CORE). All-Cause Hospital-Wide Measure Updates and Specifications Report Hospital-Level 30-Day Risk-Standardized Readmission Measure – Version 6.0. Prepared For Centers for Medicare & Medicaid Services (CMS) March 2017. https://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetTier4&cid=1219069855841. Published 2017. Accessed 24 January, 2019.
35. Ford DW, Goodwin AJ, Simpson AN, Johnson E, Nadig N, Simpson KN. A Severe Sepsis Mortality Prediction Model and Score for Use With Administrative Data. Crit Care Med. 2016;44(2):319–327. [PubMed: 26496452]
36. Lagu T, Lindenauer PK, Rothberg MB, Nathanson BH, Pekow PS, Steingrub JS, Higgins TL. Development and validation of a model that uses enhanced administrative data to predict mortality in patients with sepsis. Crit Care Med. 2011;39(11):2425–2430. [PubMed: 22005222]

37. Schwarzkopf D, Fleischmann-Struzek C, Ruddel H, Reinhart K, Thomas-Ruddel DO. A risk-model for hospital mortality among patients with severe sepsis or septic shock based on German national administrative claims data. PLoS One. 2018;13(3):e0194371. [PubMed: 29558486]

38. Simpson AN, Wilmskoetter J, Hong I, Li CY, Jauch EC, Bonilha HS, Anderson K, Harvey J, Simpson KN. Stroke Administrative Severity Index: using administrative data for 30-day poststroke outcomes prediction. J Comp Eff Res. 2018;7(4):293–304. [PubMed: 29057660]

Key Points

- We developed 7 markers of disease severity that can be broadly applied to claims data.
- For selected conditions, these markers improved prediction of outcomes.
- However, when applied broadly to the measurement of multimorbidity, disease severity markers did not improve prediction of outcomes.

Why Does This Matter?

- Claims-based measures of multimorbidity have many valuable uses; the failure of disease severity markers to improve multimorbidity assessment shows that alternate methods will be required to better measure multimorbidity in older adults.

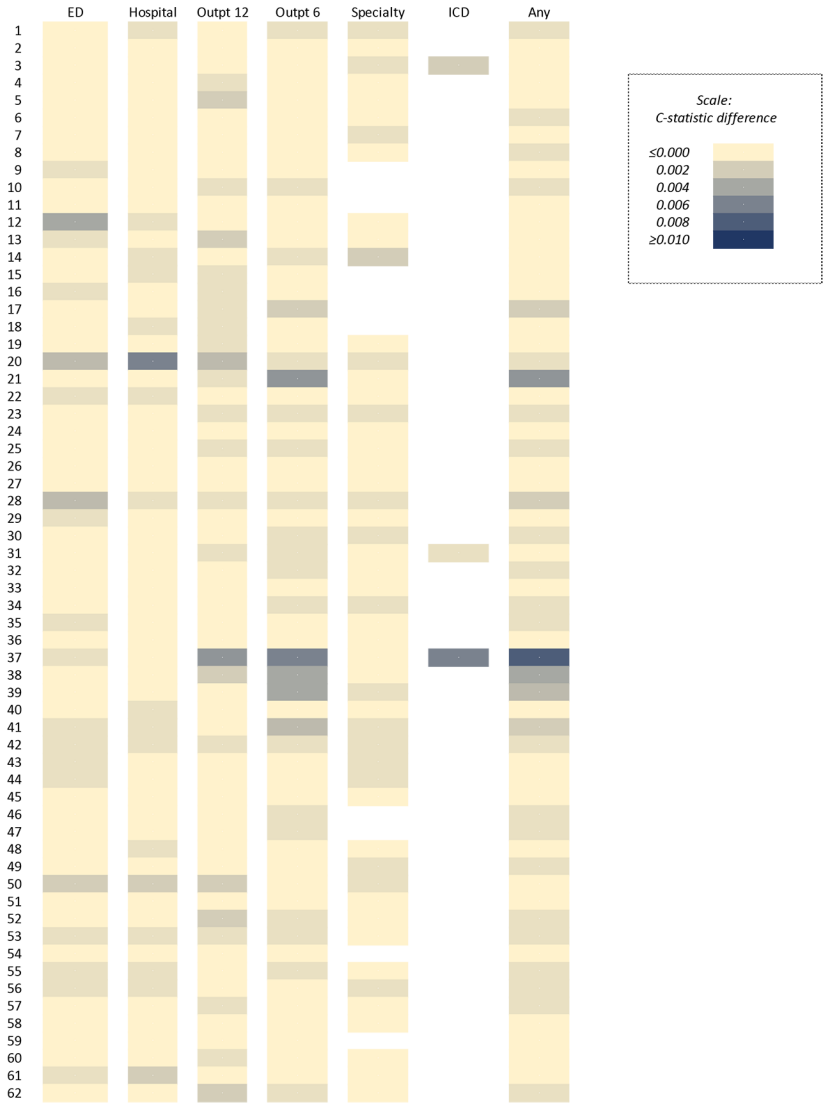

Figure 1. Change in c-statistics from incorporating markers of disease severity, for 7 disease severity markers across 62 candidate conditions – ADL decline outcome

Columns correspond to each of 7 markers of disease severity: (1) Emergency department visit for condition in the past year ("ED"), (2) Hospitalization for condition in past year ("Hospital"), (3) ≥ 4 outpatient visits for condition in the past 12 months ("Outpatient 12"), (4) ≥1 outpatient visit for condition in past 6 months ("Outpatient 6"), (5) Outpatient visit to a specialist for that condition in past 6 months ("Specialty"), (6) ICD9 code directly designating disease severity ("ICD," only available for 4 of 62 conditions), (7) severe disease on any of the above ("Any").

The y-axis represents each of the 62 conditions shown in Table 2. Each cell is shaded proportional to the gain in c-statistic for predicting future ADL decline obtained by using that marker of disease severity compared to using only the presence or absence of disease; darker shades correspond to a larger gain in c-statistic. Specifically, data represent the differences in c-statistics between a model containing age, sex, and the single disease of interest marked present vs absent, and a model containing age, sex, and the single disease of

interest represented as separate indicator variables for “disease-mild” and “disease-severe.” A shade of yellow corresponds to a difference of <=0.000; darker shades correspond to larger differences, with the largest difference (medium-dark blue) shown on this figure being 0.008. A shade of white corresponds to the severity marker not being applicable to that condition. See Supplementary Figures 6A and 6B for a more granular representation of these data.

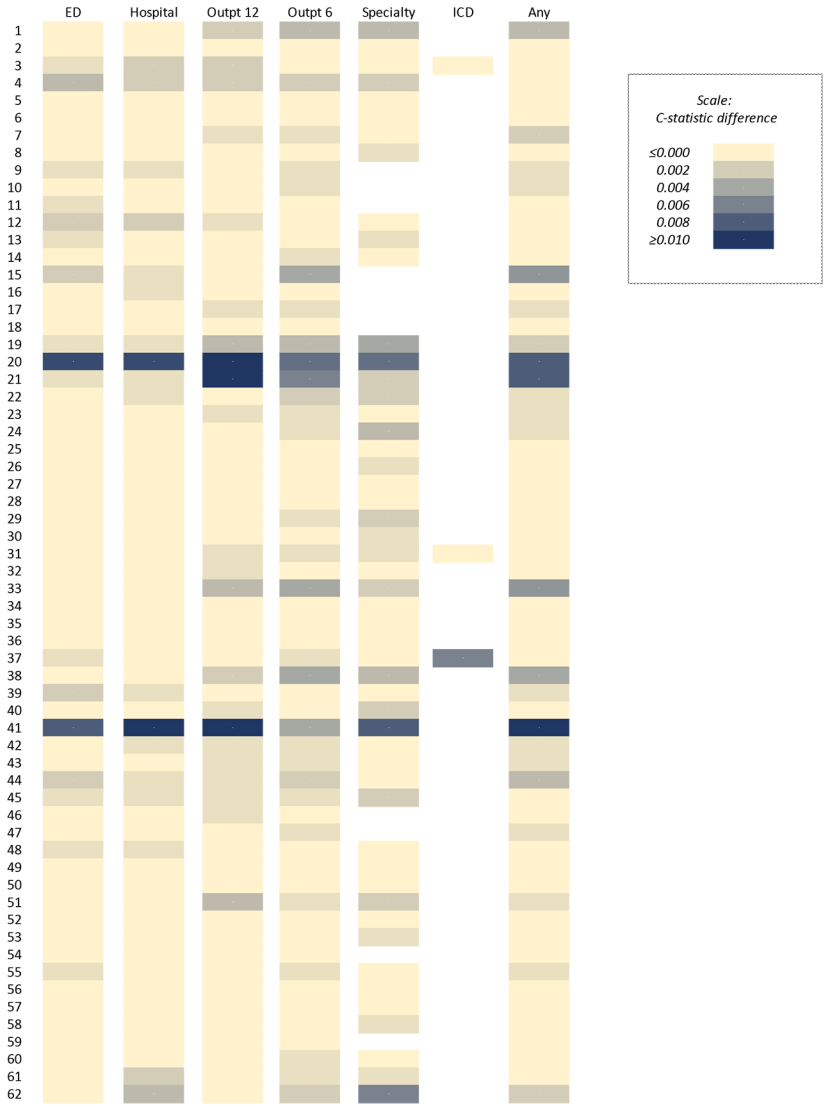

Figure 2. Change in c-statistics from incorporating markers of disease severity, for 7 disease severity markers across 62 candidate conditions - Death outcome

Columns correspond to each of 7 markers of disease severity: (1) Emergency department visit for condition in the past year ("ED"), (2) Hospitalization for condition in past year ("Hospital"), (3) ≥ 4 outpatient visits for condition in the past 12 months ("Outpatient 12"), (4) ≥1 outpatient visit for condition in past 6 months ("Outpatient 6"), (5) Outpatient visit to a specialist for that condition in past 6 months ("Specialty"), (6) ICD9 code directly designating disease severity ("ICD," only available for 4 of 62 conditions), (7) severe disease on any of the above ("Any").

The y-axis represents each of the 62 conditions shown in Table 2. Each cell is shaded proportional to the gain in c-statistic for predicting death obtained by using that marker of disease severity compared to using only the presence or absence of disease; darker shades correspond to a larger gain in c-statistic. Specifically, data represent the differences in c-statistics between a model containing age, sex, and the single disease of interest marked present vs absent, and a model containing age, sex, and the single disease of interest

represented as separate indicator variables for “disease-mild” and “disease-severe.” A shade of yellow corresponds to a difference of <=0.000; darker shades correspond to larger differences, with differences >=0.010 shown in the darkest hue of blue. A shade of white corresponds to the severity marker not being applicable to that condition. See Supplementary Figures 6A and 6B for a more granular representation of these data.

Table 1.

Characteristics of study population

Variable	%, adjusted for survey weights *
Subject Characteristics at Study Baseline	
Age (years, median)	76
Female	58.7%
Race/ Ethnicity	
White	86.9%
African American	6.6%
Hispanic and/or Latinx	4.7%
Other	1.8%
Highest level of Education	
Less than high school diploma	19.5%
High school diploma or GED	31.3%
Some college	21.1%
College graduate and above	22.0%
Marital status	
Single or never married	4.6%
Married or partnered	57.6%
Widowed	32.0%
Divorced or separated	8.8%
Comorbid Conditions (at baseline, from ICD codes)	
Heart failure	15.9%
Chronic obstructive pulmonary disease	24.2%
Hypertension	81.5%
Diabetes	39.6%
Stroke	12.3%
Metastatic cancer	2.0 %
Cognitive impairment	11.1%
Outpatient encounters in past year	
0	2.7%
1	1.8%
≥2	95.5%
Hospitalizations in past year	
0	80.7%
1	13.0%
≥2	6.3%

Variable	%, adjusted for survey weights *
Emergency department visits in past year	
0	71.1%
1	18.2%
≥2	10.8%
Activities of daily living (ADLs) requiring dependence on another person	
0	88.7%
1	4.9%
>=2	6.3%
Instrumental activities of daily living (IADLs) requiring dependence on another person	
0	82.5%
1	9.2%
≥2	8.3%
Outcomes	
Hospitalization between baseline and 2 year follow-up *	
Alive and never hospitalized	65.7%
Alive and hospitalized at least once	25.5%
Dead	8.9%
ADL decline between baseline and 2 year follow-up *	
Alive without ADL decline	82.4%
Alive with ADL decline	8.7%
Dead	8.9%
IADL decline between baseline and 2 year follow-up *	
Alive without ADL decline	80.2%
Alive with IADL decline	10.9%
Dead	8.9%
Death by 2 years follow-up	
Alive	91.1%
Dead	8.9%

Results shown for cohort used to assess hospitalization and death (N=5102). The cohorts used to assess ADL and IADL decline were similar but excluded people who did not complete the follow-up survey, or participated in that survey but did not fully complete questions on ADL and IADL status. Results are adjusted for complex survey design including sampling weights. Outcome results were defined such that death superseded hospitalization, i.e., someone who was hospitalized and then died before 2-year follow-up was classified as dead. Hispanic and/or Latinx ethnicity were coded as a non-overlapping category with other racial groups listed.

Table 2.

Conditions tested

Condition number (as indicated on graphs)	Condition
1	Visual impairment & associated conditions
2	Auditory impairment & associated conditions
3	Ischemic heart disease (all)
4	Heart failure
5	Atrial fibrillation or flutter
6	Arrhythmias other than atrial fibrillation
7	Peripheral vascular disease
8	All chronic skin ulcers
9	Abnormal gait, difficulty walking
10	Falls
11	Weakness, muscle weakness
12	Hypertension
13	Delirium
14	Syncope
15	Chronic malaise or fatigue
16	Chronic pain including fibromyalgia
17	Obesity
18	Weight loss / malnutrition / cachexia; adult failure to thrive; or debility
19	Chronic obstructive pulmonary disease (COPD)
20	Respiratory disorders other than asthma/COPD
21	Lung disease due to external agents, fibrotic lung disease, or related
22	Upper gastrointestinal disease (ulcer, gastritis, duodenitis, dyspepsia, reflux)
23	Crohn’s disease & ulcerative colitis
24	Diverticulosis & diverticulitis
25	Esophageal diseases (specified)
26	Esophageal disorders (not otherwise specified)
27	Dysphagia
28	Gallbladder disease
29	Constipation
30	Fecal incontinence
31	Liver disease (not otherwise specified)
32	Benign prostatic hypertrophy / lower urinary tract symptoms (BPH/LUTS)
33	Chronic renal insufficiency
34	Female pelvic organ prolapse
35	Other and unspecified female genital disorders
36	Female genital and reproductive tract disorders

Condition number (as indicated on graphs)	Condition
37	Diabetes
38	Hyperlipidemia
39	Fluid and electrolyte disorders
40	Metastatic cancer (solid tumors)
41	Non-metastatic cancer (solid tumors)
42	Hematologic malignancy (acute or intermediate)
43	Anemia, iron deficiency
44	Anemia, other or unspecified
45	Venous thromboembolic disease
46	HIV infection
47	Hepatitis
48	Cognitive impairment
49	Cerebrovascular disease or paralysis
50	Epilepsy
51	Parkinson’s disease
52	Hereditary and degenerative nervous system conditions other than Parkinson’s disease
53	Neuropathies (peripheral neuropathy)
54	Tobacco use
55	Alcohol abuse
56	Depression
57	Anxiety disorder
58	Psychoses
59	Insomnia and sleep disorders other than apnea
60	Hip, long bone, or pathological fracture
61	Back pain and related disorders
62	Osteoarthritis and other non-traumatic joint disorders, not including rheumatoid arthritis and related diseases (any location)

Table 3:

Difference in c statistics between multimorbidity models that do vs. do not account for disease severity.

		Outcome							
		ADL decline		IADL decline		Hospitalization		Death	
		Model c-statistic	Gain in c-statistic compared with original model	Model c-statistic	Gain in c-statistic compared with original model	Model c-statistic	Gain in c-statistic compared with original model	Model c-statistic	Gain in c-statistic compared with original model
Model ***without*** disease severity markers	Original model	0.790	--	0.755	--	0.711	--	0.813	--
Models ***with*** disease severity markers	Re-derived Model 1	0.787	−0.003	0.757	0.002	0.711	<0.001	0.815	0.002
	Re-derived Model 2	0.789	−0.001	0.756	0.001	0.710	−0.001	0.818	0.005
	Appended Model 1	0.794	0.004	0.761	0.006	0.713	0.002	0.817	0.004
	Appended Model 2	0.790	<0.001	0.755	<0.001	0.711	<0.001	0.813	<0.001

Original model: Full multimorbidity index model that includes age, sex, and indicators corresponding to the presence vs absence of 17–26 chronic conditions (without accounting for disease severity). Separate models were available for each outcome.

Re-derived model 1: The full multimorbidity index model was re-derived using indicators of disease severity applied to the 62 conditions of interest. Separate models were derived for each outcome.

Re-derived model 2: The full multimorbidity index was re-derived using a mix of disease present vs absent indicators and indicators for disease severity applied to the 62 conditions of interest. Separate models were derived for each outcome.

Appended model 1: Disease severity markers that added gain in c-statistic of >0.002 in single-disease models were appended to the variable list in the original model.

Appended model 2: A summary variable representing the number of conditions considered “severe” for that subject, among the 62 candidate conditions, was appended to the completed original model.

For all analyses shown in this table, severe disease was defined as a condition being coded as the primary reason for visit in >=4 outpatient encounters in the previous 12 months.

In some cases, rounding of results had made figures in the “gain in c-statistic” column not exactly equal to the difference between the corresponding original and revised model.