

Draft Manuscript -- Please Do Not Distribute

Medicare Advantage Star Plan Ratings Are Not Associated with Patient Health Outcomes

Eva H. DuGoff, PhD, MPP

Assistant Professor
Department of Population Health Sciences
School of Medicine and Public Health
University of Wisconsin-Madison,
Madison, WI

Abstract: 173
Word Count: 3,492 words
Exhibits: 4

Financial Disclosures: This study was funded by an AcademyHealth New Investigator Award (PI: DuGoff).

Abstract

Centers for Medicare and Medicaid Services (CMS) rates Medicare Advantage plan quality using a Star Rating using measures of clinical quality, patient experience, and plan infrastructure. While initially created to support consumer decision-making, the Star Rating is also used as the basis for determining pay for performance bonuses and contract terminations. While offering a mechanism for consumers and program administrators to discriminate between health plans, we know little about whether the Medicare's Star Rating is correlated with patient-reported health. We find that the Star Rating is not associated with the probability that a patient maintains health over two years. Analyses examining plan performance by domains found plan performance on a composite of access, process, and outcome measures were positively related to patient health, while plan performance on patient experience and intermediate were negatively associated with patient health. Based on these findings, we recommend that CMS re-consider the measures included in the Star Rating index and the weighting strategy to align the Star Rating with patient reported health.

Introduction

Comparing health plans is a difficult task, and may be more demanding for older adults who face a large number of plan options through the Medicare Advantage (MA) program. [1] To support informed consumer choices and incentivize quality improvements in the MA program, Centers for Medicare and Medicaid Services (CMS) created the Star Ratings, a composite measure assessing plan quality and performance. Today the Star Rating measure is featured on the Medicare plan finder website for consumers and integrated into the program's administration. Star ratings are used to determine pay for performance payments, opportunities to enroll consumers, and, in 2016, contract terminations.[2, 3] While offering a mechanism for consumers and program administrators to discriminate between health plans, we know little about the validity of Medicare's Star Rating. In this paper, we focus on the concurrent validity by examining the association MA plans' Star Ratings on changes in patient-reported health status.

Validity is defined as whether a tool (e.g., a performance rating) is "measuring what is intended." [4] There are a number of types of validity that are commonly assessed that range in rigor. For example, at the face validity is considered to be less rigorous because it relies on expert evaluation of whether a tool "appears" to be valid. While criterion validity, which generally uses a gold standard to assess if a tool is valid. Because there is no gold standard for health plan quality, we focus on concurrent validity, which examines the relationship between the tool and other accepted measures of quality. We focus on patient-reported health outcomes because it is generally agreed that health plans are accountable for patient health and it is an important outcome to patient.[5-7]

We link publicly reported overall star rating data to patient health outcome survey data to examine the association between MA Star Rating and changes in patient health outcomes over a two-year period. Because previous research has found that the presence of multiple chronic conditions is related with worse self-reported health, we examine for differences in the relationship between plan performance and patient reported health by baseline comorbidities using an interaction term. The results of this study will provide additional information to consumers on the applicability of the MA Star Rating to their health.

History of Medicare Advantage Star Rating

The MA overall star rating is an index measure rating plans on a scale from 1 star (low) to 5 stars (high). The ratings are updated every fall and spring and published on the Medicare plan finder website. In 2008, the Star Rating was based on clinical performance data collected by the Healthcare Effectiveness Data Information System (HEDIS), Consumer Assessment of Healthcare Providers and Systems (CAHPS), and surveillance data collected by CMS. At that time Star Ratings were reported separately for Part C (Medicare A and B covered services) and Part D (prescription drugs). Since then, the Star Rating methodology has continued to evolve incorporating additional clinical quality measures, weights, and patient survey measures from the Medicare Health Outcomes Survey (HOS). In 2011, CMS reported an overall Star Rating for Medicare Advantage plans combining Part C and Part D performance measures.

In 2014, the Overall Star Rating was the weighted average of 48 unique measures drawn from patient surveys (HOS and CAHPS), clinical process measures (HEDIS), and other administrative data collected by CMS. In order to receive an Overall Star Rating, MA plans were required to have data available on at least half of the underlying measures, a Part C summary score, and a Part D summary score. Individual measures are weighted according to its category: outcome and intermediate outcome measures receive a weight of 3; patient experience measures receive a weight of 1.5; and, process measures receive a weight of 1. The overall star rating is a weighted average of the star ratings for the available underlying measures and an integration factor to reward plans with consistently high scores on underlying measures.

The Patient Protection and Affordable Care Act of 2010 required CMS to use the Star Rating to reward MA plans receiving 4 or more stars with bonus payments.[8] Under its demonstration authority, however, CMS expanded the pay for performance initiative to include 3 and 3.5 star plans and offer larger bonus payments. Under a 3-year demonstration program, CMS scaled such that the highest rated plans received greatest incentive and 3 and 3.5 star plans received smaller bonus payments.[2] According to CMS, the goal of the demonstration was to provide additional incentives for improvement to 3 and 3.5 star plans as well as to test whether incentive payments would increase the number of highly rated plans. The program came under some criticism from Medicare Payment Advisory Commission and Government Accountability Office for its substantially high cost (\$8 billion over three years).[9, 10] Following the conclusion of the demonstration, CMS has continued to use the Star Rating to offer bonus payments to plans receiving 4 or more stars. In 2016, CMS has also said it will terminate MA plan contracts if a plan has received less than 3 stars on its Part C summary score for the past three years.[3]

In the literature, there is some evidence that first-time consumers are more likely to choose a highly rated plans.[11] The Kaiser Family Foundation has also reported that enrollment in MA plans with 4 and 4.5 stars has increased from 37% to 61% from 2013 to 2015.[12] However, enrollment in the highest rated plans has been stable at 8% over this same time period. One reason enrollment in the highest rated plans has not substantially increased is that older adults have reported some skepticism about the value of the Star Rating.[1] A focus group study found that some older adults dismissed the Star Rating because they did not understand what the score represented and whether it was consistent with their values.

In recent years, the Star Rating has come under some criticism. Some researchers and health plans have raised concerns that the Star Rating disproportionately rates plans serving individuals dually eligible for Medicare and Medicaid.[13, 14] Relatedly, one study did find that enrollee characteristics did influence Medicare Part D plan performance scores.[15] In response, CMS has commissioned research from RAND to better understand the influence of patient-level characteristics on plan performance scores. Congress, through the IMPACT Act, has also requested the Secretary of Health and Human Services study of these issues and report to Congress in 2016.

Performance Indices and Health Outcomes

Studies of performance indices have largely focused on hospital performance [16-20] and nursing homes [21, 22]. These studies have reported mixed results. For example, studies have found *US News & World Report* hospital rankings are correlated with 30-day mortality rates [16, 17], but not hospital readmissions [23]. A recent report also documented substantial disagreements between four publicly available hospital ratings systems [18]. The authors found that only 10 percent of hospitals rated as a high performer on one rating system was also rated as a high performer by another. The disagreements between rating systems is perhaps not surprising given the heterogeneity in methodological approaches and underlying data.

We are not aware of any studies that have directly examined the relationship between MA Star Ratings and patient health outcomes. Two previous studies have examined the relationship between health plan performance measures and patient reported health outcomes in MA. [24, 25] A study by Harman and colleagues found that plan performance on composite index of diabetes processes of care was associated with greater probability of good mental health after two years and a composite index of diabetes intermediate health measures was positively associated with good physical health. Similarly, Lied and colleagues found that patient self-reported health was positively correlated with process of care measures.

Methods

Data. This study uses the Cohort 14 Medicare Health Outcomes Survey (HOS) fielded by mail and telephone in 2011 and 2013. We use publicly reported health plan performance data published in 2014 because it draws on measures largely collected in 2012 – the intervening year between the baseline and follow up HOS surveys.

HOS is a longitudinal survey of Medicare Advantage enrollees occurring at two-year intervals. The HOS uses an overlapping panel design such that in each year one cohort is completing the baseline survey and another cohort is completing the follow up survey. The HOS collects individual demographic, preventive care, and health status information from a random sample of members from each MA contract. The HOS collects several health status measures including two generic health related quality of life measures: CDC's number of unhealthy days, Veterans Rand-12 which estimates an individuals physical and mental component summary score. We use focus on the Cohort 14 data because these data are the most recent release providing both baseline and follow up data.

To measure health plan performance, we use the publicly available 5 star data published on the CMS website. The Star Rating is a composite measure aggregating 48 different performance indicators. The underlying measures assessing outcomes, intermediate outcomes, patient experience, access, and processes of care.

Study Sample. Our study sample includes non-institutionalized older adults enrolled in Medicare Advantage plans receiving a 5 star rating in 2014 who responded to the Medicare Health Outcome Survey (HOS) at baseline and follow up. In 2014, 431 Medicare Advantage plans received an overall rating by CMS of which 411 could be linked to the HOS Cohort 14

survey. A total of 103,398 HOS respondents were enrolled in one of these plans at baseline in 2011 and follow up in 2013.

HOS Cohort 14 baseline survey included 577,874 MA enrollees. A total of 297,570 individuals responded to the baseline survey (response rate 51.5%) of which 252,580 were age 65 or older.

Health Plan Definition. We define a health plan or MA organization by its contract number. Contracts may include multiple smaller units (plans). The contract is the appropriate unit of analysis because the MA Star Rating is assigned by MA organization contract.

Covariates. Based on the Aday-Andersen framework, we posit that patient-level health outcomes may vary based on a person's predisposing, enabling, and need factors.[26] We control for individual age, sex, self-reported race (Black, Asian, and Hispanic), marital status (married vs not), home ownership (own home vs not), Medicaid status (yes vs no), number of chronic conditions (0-1, 2-5, 6+), and deficiencies in activities of daily living (ADLs; 0,1,2+).

We expect that health plan characteristics may also influence individual selection into a health plan. We control for health plan tax status (for profit and non-profit), enrolled population (<500,500-1200,1200-3000,3000-5000, and 10,000+), type of plan (HMO, PPO, and private fee for service), average proportion of enrollees with Medicaid, and average proportion of enrollees graduating high school.

Health Plan Performance Measures. We define health plan performance using two approaches. We use the overall Star Rating measure that combines a plan's Part C and Part D performance published by CMS in 2014. The Overall Star Rating ranges from 2 to 5 stars. In 2014, relatively few plans received the highest (11 plans received 5 stars) or lowest (1 plans received 2 stars) scores. Thirty-four percent 233 plans did not receive an Overall Star Rating due to insufficient data (58 plans) and time of entry into the program (175 plans were too new to be measured).

CMS categories performance measures underlying the star rating into five types: access, process, patient experience, intermediate outcomes, and outcomes. In order to better understand the independent association of each of these categories on patient reported health, we calculated category-specific performance scores. Consistent with previous studies we used a composite score: the average proportion of eligible enrollees meeting measure criteria on all underlying measures. [25] We reverse coded measures where plan performance is defined as the proportion of enrollees out of compliance (e.g., 30-day hospital readmission rates) as opposed to in compliance. For the outcome category, we focus on three measures: maintaining or improving physical health, maintaining or improving mental health, and 30-day hospital readmissions. We excluded two outcome measures, C33 (Health Plan Improvement) and D07 (Drug Plan Improvement) because these measures were not reported on the same scale as other measures.

Outcome Measures. The primary dependent variable was the two-year change in the predicted probability of maintaining health using the Summary Index of Health Days. While other studies have examined good health, we focus on maintaining health because it may be a more feasible and meaningful standard for older adults, especially those with chronic disease.[25]

The Summary Index of Healthy Days is a health related quality of life instrument developed and validated by the Centers for Disease Control and Prevention.[27, 28] It is a composite of a two survey items: “Now thinking about your physical health, which includes physical illness and injury, for how many days during the past 30 days was your physical health not good?” and “Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?” The maximum number of possible unhealthy days is 30. To calculate the Summary Index of Healthy Days, we subtract 30 from the reported number of unhealthy days.

Individuals who died during the study period are missing data at follow-up. Decedents are expected to have poor health status, and therefore excluding them from the analysis would inflate the estimated 2-year change in health. We followed recommended procedures to incorporate health status changes of those who died during the study period.[25, 29, 30] First, we calculate the raw 2-year change in health as a dichotomous indicator of whether the patient’s reported number of healthy days at follow up were greater than or 1 of the reported number of healthy days at baseline. Using logistic regression, we estimate maintaining health as predicted by baseline number of Healthy Days. We then applied this model to the sample to predict the probability of maintaining health as a function of baseline Healthy Days and follow-up healthy days. We impute the probability of maintaining health among decedents as zero at follow-up (N=16,309, 6.5%).

Analytic Approach

We first examined the characteristics of respondents and non-respondents in 2011 to assess for differences in survey response. We calculated survey weights accounting for the different probabilities of selection based on plan size, response at baseline, and response at follow up. We used raking procedure to calibrate weights to age and race characteristics of the survey sample. Survey weights were trimmed at the 95th percentile to reduce the influence of outliers.

We used multivariate linear regression models to model the relationship between plan performance and patient health. We accounted for the stratified sampling design and survey weights using the survey package in R. We ran models assessing the overall association of Star Rating on patient health and then examined whether the relationship between Star Rating and patient health varied by level of baseline comorbidities.

In sensitivity analyses, we tested the robustness of our results to the definition of patient reported health. We examined models where the dependent variable was defined as using the VR-12 physical component score and mental component score. We also examined the sensitivity of the results to models where MA plan Contract was included as a fixed effect.

Analyses were conducted using RStudio using the survey package. [31, 32] This study was determined to be exempt from review by the University of Wisconsin-Madison Institutional Review Board.

Results

We linked 130,939 HOS survey respondents representing 11 million MA enrollees to 411 Medicare Advantage plans with a Star Rating (Exhibit 2). The majority of MA enrollees were 65 years or older, female, high school graduates, and home owners. Twenty-nine percent of the sample reported living with two or more deficiencies in activities of daily living, and 78% reported living with two or more chronic conditions. Of baseline respondents, 11% died prior to the follow-up survey.

MA Star Rating was not associated with 2-year changes in patient-reported health (Exhibit 3). As expected, the probability of maintaining health was negatively associated with the presence of multimorbidity and deficiencies in activities of daily living. We found that older adults with multiple chronic conditions at baseline are less likely to maintain their health: 2-5 chronic conditions Beta: -0.6 (95%CI -1.0, -0.3) and older adults with six or more chronic conditions Beta -3.9 (95%CI -4.5, -3.2). Maintaining health was less likely in MA enrollees over 70 and those enrolled in Medicaid. Home ownership, being married, and a high school education were positively related with maintaining health. We did not find that the relationship between multimorbidity mediates the relationship between plan performance and patient reported health.

The relationship between MA plan performance on access, process, patient experience, and outcomes composite indices and maintaining health varied (Exhibit 4). MA plan performance on composite indices of access (0.4, 95%CI: 0.2, 0.7), processes (0.5, 95%CI: 0.0, 1.1), and outcomes (0.6, 95%CI: 0.1, 1.1) were positively related to maintaining health. Plan performance on patient experience (-1.9, 95%CI: (-3.0, -0.8) and intermediate outcomes (-0.9, 95%CI: -1.4, -0.4) were negatively associated with patient health. Deficiencies in ADLs and multimorbidity were negatively associated with patient reported health. Indicators of higher socio-economic status were associated with greater probability of maintaining health. Enrollee multimorbidity level did not mediate the relationship between plan performance and patient reported health.

Discussion

Medicare Advantage Star Ratings were not associated with two-year changes in the probability of maintaining health. Plan performance on access to care, process, and outcome measures were associated with greater likelihood of maintaining health. Patient experience and intermediate outcomes were associated with lower likelihood of maintaining health.

Overall, these results are consistent with other studies that find that particular performance domains are related to patient health outcomes. Lied and colleagues found process measures to be associated with patient self-reported health.[24] Harman and colleagues found that diabetes process measures were associated with better patient

reported health after two years.[25] This study extends this literature by examining MA plan performance across all enrollees using plan performance measures currently used by the MA Quality Bonus Program.

Aggregate plan performance measures could provide crucial information to Medicare beneficiaries looking to enroll in a private Medicare Advantage plans. However, the results of this study indicate that the Star Rating on aggregate may not indicate plans that are able to achieve better health outcomes for their enrollees. CMS administrators should consider reviewing the measures included in the Star Rating to focus not on performance measures that can be feasibly collected in the field, but also the measures that are predictive of better patient outcomes.

An advantage of an index is that it pools together multiple measures increasing overall reliability. However, in creating an index, it is important to consider whether the quality measures represents a type of care or construct or if the measure is relevant only to a specific subgroups. It isn't the case that a single index will be applicable to all individuals and the applicability may change over time. These are issues that should be regularly monitored by CMS and the Star Rating should be revised as necessary.

As expected, outcome scores were associated with positive patient health outcomes overall. Higher scores in these domains largely benefited individuals with better health status at baseline. It is unclear if patients directly benefit from greater adherence to the underlying performance measures themselves, or if these domain scores are capturing a gestalt within the health plan that may have a direct or indirect benefit for patients such as greater contact with the health care system and opportunities to discuss one's care management.

We were surprised to find that higher patient experience and intermediate scores were associated with poorer outcomes for older adults. This study does not investigate why these measures may be associated with poor patient health outcomes such as the health outcomes of the relevant patient populations.

Performance measurement faces many challenges. It is thought that quality measures should have some relationship with health outcomes to be valid. Quality measures have come under some scrutiny for issues with face validity. An advantage of an aggregate ranking measure (or index) is that it doesn't rely on a single measure and may be able to identify more substantial variations between plan performance.

The findings of this study should be considered in light of its limitations. Survey data is subject to non-response and respondent bias. We examine overall plan performance and performance along commonly recognized measurement domains, but we cannot identify which measures are most closely associated with better outcomes among older adults. We use the most recent data available, however, these data may be influenced by the introduction of the MA pay for performance program—Quality Bonus Program—in 2012.

In summary, novel data linking health plan performance to patient-level health outcomes we find that MA Star Ratings are not positively associated with patient health outcomes. As

Draft Manuscript -- Please Do Not Distribute

value based payment models spread throughout the Medicare program, one way to assess the validity of these measures is to examine the relationship between the aggregate score and changes in patient health.

Exhibit 1. Medicare Advantage Plan Performance Overall and by Domains (N = 411), 2014

	Number of Items	Exemplar Measure	Mean Score (SE)
Star Rating	48		3.6 (0.6)
Access	7	Beneficiary Access and Performance Problems	84.5 (8.3)
Process	18	Breast Cancer Screening	69.2 (5.1)
Patient Experience	10	Rating of Health Care Quality	88.4 (2.5)
Intermediate Outcomes	8	Diabetes Care – Cholesterol Controlled	53.8 (5.5)
Outcomes	3	Plan All-Cause Readmissions	78.7 (3.8)

Exhibit 2. Survey Weighted Sample Characteristics, 2011

	2011
N	11,043,344
Age (%)	
<65	15.3
65-69	25.2
70-74	21.3
75-79	16.4
80-84	11.6
85+	10.2
Female (%)	56.6
Black	10
Hispanic	12
Medicaid (%)	
Medicaid	19.3
Missing	0.5
Education (%)	
HS Grad	75.9
Missing	2.3
Home Ownership (%)	
Own Home	66
Missing	2.8
Married (%)	
Married	53
Missing	2
ADL (%)	
1	12.7
2+	29
Chronic Conditions (%)	
Low	22.8
Middle	61.4
High	15.8
Death	11

Exhibit 3. Plan Performance on Star Rating and 2-year Change in the Probability of Maintaining Health

	Beta	95% CI
Overall	0.011	(-0.329, 0.352)
plantaxNon-Profit	0.315	(-0.097, 0.728)
plantypePPO	-0.758***	(-1.164, -0.352)
plantypeOther	0.579	(-0.248, 1.407)
mcaid.gmc	-6.507***	(-7.291, -5.724)
agecat_18.cwc	4.290***	(3.683, 4.898)
agecat_70.cwc	-1.165***	(-1.568, -0.763)
agecat_75.cwc	-2.503***	(-2.999, -2.007)
agecat_80.cwc	-5.841***	(-6.534, -5.147)
agecat_85.cwc	-13.698***	(-14.635, -12.760)
female.cwc	3.077***	(2.711, 3.443)
black.cwc	0.401	(-0.288, 1.090)
hispanic.cwc	2.463***	(1.677, 3.248)
medicaid.cwc	-3.093***	(-3.846, -2.341)
adl_1.cwc	-3.365***	(-3.944, -2.785)
adl_2.cwc	-9.276***	(-9.817, -8.736)
cc_middle.cwc	-0.644***	(-1.005, -0.283)
cc_high.cwc	-3.906***	(-4.583, -3.230)
home.cwc	2.267***	(1.776, 2.757)
hs.cwc	1.674***	(1.138, 2.210)
married.cwc	0.750***	(0.335, 1.166)

Exhibit 4. Plan Performance on Access, Process, Patient Experience, Intermediate, Outcomes and 2-year Change in the Probability of Maintaining Health

	Beta	95% CI
access	0.440***	(0.206, 0.673)
process	0.531**	(0.013, 1.049)
ptexp	-1.906***	(-3.026, -0.786)
int	-0.898***	(-1.426, -0.369)
outcomes	0.610**	(0.089, 1.131)
plantaxNon-Profit	0.470**	(0.046, 0.895)
plantypePPO	-1.077***	(-1.501, -0.654)
plantypeOther	0.767*	(-0.079, 1.613)
mcaid.gmc	-7.065***	(-7.824, -6.306)
agecat_18.cwc	4.249***	(3.642, 4.857)
agecat_70.cwc	-1.160***	(-1.562, -0.757)
agecat_75.cwc	-2.515***	(-3.011, -2.019)
agecat_80.cwc	-5.857***	(-6.550, -5.163)
agecat_85.cwc	-13.716***	(-14.654, -12.778)
female.cwc	3.072***	(2.706, 3.438)
black.cwc	0.379	(-0.309, 1.068)
hispanic.cwc	2.456***	(1.670, 3.241)
medicaid.cwc	-3.096***	(-3.849, -2.344)
adl_1.cwc	-3.363***	(-3.943, -2.783)
adl_2.cwc	-9.278***	(-9.818, -8.738)
cc_middle.cwc	-0.646***	(-1.007, -0.285)
cc_high.cwc	-3.907***	(-4.583, -3.230)
home.cwc	2.260***	(1.770, 2.750)
hs.cwc	1.674***	(1.138, 2.210)
married.cwc	0.741***	(0.325, 1.156)

References

1. Jacobson, G., et al., *How are Seniors Choosing and Changing Health Insurance Plans?*. 2014: Washington, DC. p. 25.
2. *Proposed Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs for Contract Year 2012 and Demonstration on Quality Bonus Payments*, D.o.H.a.H. Services, Editor. 2011, Centers for Medicare and Medicaid Services: Washington, DC. p. 10.
3. *Announcement of Calendar Year (CY) 2015 Medicare Advantage Capitation Rates and Medicare Advantage and Part D Payment Policies and Final Call Letter*, D.o.H.a.H. Services, Editor. 2015, Centers for Medicare and Medicaid Services: Baltimore, MD. p. 154.
4. Streiner, D.L. and G.R. Norman, *Health measurement scales : a practical guide to their development and use*. 4th ed. 2008, Oxford ; New York: Oxford University Press. xvii, 431 p.
5. Nerenz, D.R., *Accountability for health outcomes and the proper unit of analysis: what do the experts think?* Int J Qual Health Care, 1998. **10**(6): p. 539-46.
6. Black, N., *Patient reported outcome measures could help transform healthcare*. Bmj, 2013. **346**: p. f167.
7. Murphy, M., et al., *Patient and practitioners' views on the most important outcomes arising from primary care consultations: a qualitative study*. BMC Fam Pract, 2015. **16**: p. 108.
8. Jacobson, G., et al., *Medicare Advantage Plan Star Ratings and Bonus Payments in 2012* 2011, Kaiser Family Foundation: Washington, DC. p. 17.
9. Gibson, L.H., *Medicare Advantage Quality Bonus Payment Demonstration*, U.G.A. Office, Editor. 2012: Washington, DC. p. 10.
10. Hackbarth, G.M., *MedPAC comment on CMS's proposed rule entitled Medicare Program; Proposed Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs for Contract Year 2012 and Other Proposed Changes*, M.P.A. Commission, Editor. 2011: Washington, DC. p. 7.
11. Reid, R.O., et al., *Association between medicare advantage plan star ratings and enrollment*. JAMA, 2013. **309**(3): p. 267-274.
12. Jacobson, G., et al., *Medicare Advantage 2015 Spotlight: Enrollment Market Update*, in *Issue Brief*. 2015.
13. Inovalon, *The Impact of Dual Eligible Populations on CMS Five-Star Quality Measures and Member Outcomes in Medicare Advantage Health Plans*, in *Inovalon Research Brief*. 2013, Inovalon, Inc.: Bowie, MD. p. 62.
14. Association for Community and Affiliated Plans, *Medicare's Quality Incentive System Does Not Adequately Account for Special Needs of Dual-Eligible Populations*. 2012, Association for Community and Affiliated Plans: Washington, DC. p. 11.
15. Young, G.J., et al., *Socioeconomic characteristics of enrollees appear to influence performance scores for medicare part D contractors*. Health Aff (Millwood), 2014. **33**(1): p. 140-6.
16. Wang, O.J., et al., *"America's Best Hospitals" in the treatment of acute myocardial infarction*. Arch Intern Med, 2007. **167**(13): p. 1345-51.

17. Chen, J., et al., *Do "America's Best Hospitals" perform better for acute myocardial infarction?* N Engl J Med, 1999. **340**(4): p. 286-92.
18. Austin, J.M., et al., *National hospital ratings systems share few common scores and may generate confusion instead of clarity.* Health Aff (Millwood), 2015. **34**(3): p. 423-30.
19. Krumholz, H.M., et al., *Evaluation of a consumer-oriented internet health care report card: The risk of quality ratings based on mortality data.* JAMA, 2002. **287**(10): p. 1277-1287.
20. Chen, J., et al., *Performance of the '100 top hospitals': what does the report card report?* Health Aff (Millwood), 1999. **18**(4): p. 53-68.
21. Calikoglu, S., C.S. Christmyer, and B.U. Kozlowski, *My eyes, your eyes--the relationship between CMS five-star rating of nursing homes and family rating of experience of care in Maryland.* J Healthc Qual, 2012. **34**(6): p. 5-12.
22. Williams, A., J.K. Straker, and R. Applebaum, *The Nursing Home Five Star Rating: How Does It Compare to Resident and Family Views of Care?* Gerontologist, 2014.
23. Mulvey, G.K., et al., *Mortality and readmission for patients with heart failure among U.S. News & World Report's top heart hospitals.* Circ Cardiovasc Qual Outcomes, 2009. **2**(6): p. 558-65.
24. Lied, T.R. and S.H. Sheingold, *Relationships Among Performance Measures for Medicare Managed Care Plans.* Health Care Financ Rev, 2001. **22**(3): p. 23-33.
25. Harman, J.S., et al., *Association of Health Plans' Healthcare Effectiveness Data and Information Set (HEDIS) performance with outcomes of enrollees with diabetes.* Med Care, 2010. **48**(3): p. 217-23.
26. Aday, L.A. and R. Andersen, *A framework for the study of access to medical care.* Health Serv Res, 1974. **9**(3): p. 208-20.
27. Zullig, K.J., *Creating and using the CDC HRQOL healthy days index with fixed option survey responses.* Qual Life Res, 2010. **19**(3): p. 413-24.
28. Moriarty, D.G., M.M. Zack, and R. Kobau, *The Centers for Disease Control and Prevention's Healthy Days Measures - population tracking of perceived physical and mental health over time.* Health Qual Life Outcomes, 2003. **1**: p. 37.
29. Diehr, P., et al., *Accounting for deaths in longitudinal studies using the SF-36: the performance of the Physical Component Scale of the Short Form 36-item health survey and the PCTD.* Med Care, 2003. **41**(9): p. 1065-73.
30. Diehr, P., et al., *Transforming self-rated health and the SF-36 scales to include death and improve interpretability.* Med Care, 2001. **39**(7): p. 670-80.
31. Team, R., *RStudio: Integrated Development for R.* RStudio, Inc., Boston, MA. URL <http://www.RStudio.com/ide>, 2014.
32. Lumley, T., *Survey: Analysis of Complex Survey Samples.* 2011, R package version 3.26.