

# UC Irvine

## UC Irvine Previously Published Works

### Title

Using contingent valuation to develop consumer-based weights for health quality report cards.

### Permalink

<https://escholarship.org/uc/item/7gv2367s>

### Journal

Health services research, 54(4)

### ISSN

0017-9124

### Authors

Weimer, David L  
Saliba, Debra  
Ladd, Heather  
et al.

### Publication Date

2019-08-01





### DOI

10.1111/1475-6773.13155

Peer reviewed

## RESEARCH ARTICLE

# Using contingent valuation to develop consumer-based weights for health quality report cards

David L. Weimer PhD<sup>1</sup>  | Debra Saliba MD, MPH<sup>2,3,4</sup>  | Heather Ladd MS<sup>5</sup>  |  
Yuxi Shi MS<sup>5</sup> | Dana B. Mukamel PhD<sup>5</sup> 

<sup>1</sup>LaFollette School of Public Affairs, University of Wisconsin – Madison, Madison, Wisconsin

<sup>2</sup>UCLA Borun Center at David Geffen School of Medicine, Los Angeles, California

<sup>3</sup>Veterans Administration GRECC, Los Angeles, California

<sup>4</sup>RAND Health, Santa Monica, California

<sup>5</sup>Department of Medicine, Division of General Internal Medicine, iTEQC Research Program, University of California, Irvine, California

## Correspondence

Dana B. Mukamel, PhD, Department of Medicine, Division of General Internal Medicine, iTEQC Research Program, University of California, Irvine, 100 Theory, Suite 120 Irvine, CA 92617-3056.  
Email: dmukamel@uci.edu

## Funding information

National Institutes of Health, Grant/Award Number: R01 AG049705

## Abstract

**Objective:** The current 5-Star composite measure for nursing homes uses expert-driven weights to combine elements of quality into a single score. We assessed the feasibility of using the contingent valuation method (CVM) to derive consumers' preference-based weights for the Nursing Home Compare report card as a potential alternative approach.

**Data Sources:** Survey of 4310 adults with nursing home experience (residents or family members of a resident) administered between September 25 and October 9, 2017.

**Study Design:** Contingent valuation method based on respondents' answers to questions about willingness-to-trade (WTT) visit travel time for better quality in seven quantitative indices included in Nursing Home Compare. We calculated WTT amounts per standard deviation change in quantitative indices to derive weights.

**Data Collection Methods:** Web-based survey.

**Principal Findings:** Contingent valuation method results are consistent with respondents making economically rational trade-offs between quality and travel time. Estimates of mean WTT vary across quantitative quality indices. They also vary in terms of respondent status and behavioral factors. Weights based on mean WTT per standard deviation vary substantially across indices, with the largest weights for inspections and staffing.

**Conclusions:** Contingent valuation method has promise as a method for deriving weights for use in summary measures that incorporate consumer preferences.

## KEYWORDS

contingent valuation method, Nursing Home Compare, quality report card, star ratings, willingness-to-trade time

## 1 | INTRODUCTION

Health care report cards present information to consumers about the quality of care offered by providers such as surgeons, hospitals, and nursing homes (NHs). To the extent that the information shifts demand to more highly ranked providers, report cards create incentives for providers to increase their quality scores. In addition,

professional pride may suffer from unfavorable comparisons to other providers.<sup>1</sup> To influence consumer choice rationally, report cards' information must be relevant and comprehensible. In order to avoid dysfunctional provider responses to report cards (eg, "teaching to the test" and "cream skimming"), reports should be based on measures that comprehensively span the important dimensions of quality and use appropriate risk adjustment. Unfortunately, these

goals of comprehensibility and comprehensiveness often conflict.<sup>2</sup> In general, the greater the number of quality dimensions covered, the less likely the report card will be comprehensible.

Nursing Home Compare (NHC), published by the Centers for Medicare & Medicaid Services (CMS), provides a large number of quantitative quality indices that pose a challenge to comprehensibility. Currently,<sup>3</sup> the quality indices reported in NHC and included in the 5-Star composite are in three categories: (a) quality measures (seven for short-stay populations and nine for long-stay care populations), (b) measures based on health inspections, and (c) facility staffing. In an effort to make these multiple quality indices more comprehensible, CMS, in 2008, introduced a star system that translates the quality indices in each area into a star rating ranging from one (lowest quality) to five stars (highest quality). An algorithm then translates the stars for these three areas into an overall facility star rating. Introduction of this star rating system had a significant impact on consumer demand.<sup>4</sup> Evidence of inflation in rating scores suggests that facilities see their scores as consequential.<sup>5</sup>

A panel of clinical and research experts provided input to CMS on the methodology for creating the star score. Specifically, the panel reviewed item performance and, based on their experience and preferences, advised CMS on the selection and relative importance to be assigned to the quality indices included in the 5-Star ratings. Currently, the overall 5-Star rating begins with the number of stars given for inspections and then adjusts for either very low or very high numbers of stars assigned for nurse staffing and 16 of the current 31 quality measures.<sup>3,6</sup>

Reliance on expert opinion to construct composite measures offers several advantages. First, experts offer familiarity and understanding of a range of care processes, structures, and quality indices. Second, experts may have information about measure reliability and validity that is relevant to selection and weighting. Third, soliciting expert opinion is relatively inexpensive. But do expert-based weights reflect the preferences of patients?

Although limited, available evidence suggests a divergence between expert-based ratings and the preferences of NH residents and family members. An analysis of the Ohio Nursing Home Family satisfaction survey found that “the star rating system does not adequately reflect consumer preferences.”<sup>7</sup> A recent California study that elicited preferences from patients and their family members at the time of hospital discharge to NHs suggests that individuals often value quality dimensions differently than the expert-based 5-Star ratings.<sup>8-10</sup> However, fully personalizing NH ratings imposes its own cognitive burdens on those selecting NHs. Are there feasible alternative ways to incorporate consumer preferences into decisions about summary scoring? A positive answer to this question is likely to have value beyond NHC as both demands for greater provider accountability and the technical capabilities for constructing report card measures increase. Indeed, CMS has implemented 5-Star rating systems for many other provider types, including health plans, hospitals, and home health agencies.

In this research, we explore the feasibility of using the contingent valuation method (CVM) to develop preference-based weights

for health care report cards. To do so, we elicit the willingness of consumers familiar with NHs' care to trade additional travel time of visitors with improvements in quality indices.

## 2 | STUDY DESIGN

We employ dichotomous choice elicitation, which is generally viewed as the CVM least susceptible to cognitive and strategic biases and therefore most reliable.<sup>11,12</sup> In common practice, a survey experiment describes a change in some good and asks respondents if they would be willing to pay some dollar amount, called a bid price, to obtain the good. If the bid prices are chosen to elicit a sufficient gradient in response rates, then the pattern of responses provides a basis for estimating respondents' mean willingness to pay for the change.

Rather than denominate bid prices in dollars, we chose to express them as additional travel times for visitors to an alternative, higher-quality, nursing home. We did this for several reasons. First, we were concerned about the challenge of explaining hypothetical out-of-pocket costs to respondents who often pay for NH care through complex insurance arrangements with varying out-of-pocket costs. Second, we anticipated trade-offs between time and quality would be less susceptible to social desirability bias because avoiding loss of visiting time was less likely to be viewed as selfish than avoiding payments. Third, unlike most applications of CVM aimed at finding a willingness-to-pay in dollars to include in cost-benefit analyses, we only require *relative* trade-offs for the construction of weights for quality indices, so that ratios of mean willingness-to-trade travel time (WTT) for quality indices would reflect their relative weights in constructing overall ratings. Although we know of no use of travel time and quality trade-offs in CVM, the commonly used time-trade-off method asks respondents to trade longevity and health utility to develop weights for quality-adjusted life years.<sup>13</sup>

As each respondent provides relatively little information in a dichotomous choice CVM, researchers generally recommend large sample sizes of over 500 respondents for each elicitation.<sup>14</sup> This requirement, combined with the necessity to limit the number of elicitations per respondent to a small number (ideally one) to avoid early elicitations priming subsequent ones, makes it potentially costly, in terms of sample size, to elicit WTT for all the currently used NHC quality indices. Therefore, in this pilot study, we elicited WTT for changes in only seven different measures with one measure repeated with a larger change as a scope test. We chose representative quality indices in each of the three categories: staffing (nursing hours per resident per day), inspections (number of citations), and quality measures. For the quality measures, we chose both long-term (activities of daily living, pain, and pressure ulcers) and short-term measures (rehospitalizations and discharge to community). The choice of the specific indices reflected consideration of inclusion of all three categories included in the current 5-Star method, the importance of the indices as markers of NH quality, and their importance to consumers.<sup>9</sup>

Table 1 summarizes the quality indices for which WTT was elicited. The first column shows the base level presented in the elicitation, the second column shows the level offered in an alternative NH that required randomly assigned additional travel time for visits, the third column displays the difference in quality, and the fourth column expresses the differences in quality as standard deviations of the quality index in national data at the time of the study.

## 2.1 | Survey data

Survey respondents were recruited from the Survey Sampling International (SSI) multimillion-member online panel, which includes over seven million panelists who agree to participate in Internet-based surveys. Comparisons between surveys based on similar samples of willing respondents made when random digit dialing telephone sampling was considered the “gold standard” mode for U.S. surveys showed similar substantive results for analyses involving survey experiments like CVM.<sup>15</sup>

The survey instrument was pretested in May 2017 through 60-minute one-on-one interviews with eight family members of current NH residents and two former (short stay) NH residents who were each paid \$125 for participation. Participants completed the online survey while sharing their computer screen and simultaneously discussing questions and responses with the survey team. The pretest contributed to improved question clarity.

In a process that preserved anonymity of respondents, SSI sent email invitations for participation to its U.S. panel. Inclusion criteria were as follows: age 18 years or older, U.S. residence, and either currently residing or having resided in a NH within the past 6 months or be closely involved with the care of a family member currently residing or having resided in a NH within the past 6 months. Participants who met the qualifying criteria were assigned to one of the eight respondent statuses based on responses to screening questions: (a) currently residing, short-term, (b) currently has a family member, short-term, (c) currently residing, long-term, (d) currently has a family member, long-term, (e) was resident in last 6 months, short-term, (f) had family member in short-term in last 6 months, (g) prior long-term

resident discharged in last 6 months, and (h) had family member in long-term in last 6 months who was discharged or died. To qualify, respondents also had to express familiarity with the quality of care in the relevant nursing home.

Qualified participants who completed the full survey were awarded SSI loyalty program points that can be exchanged online for gifts; no cash incentives were offered.

Data were collected from September 25 to October 9, 2017. Email invitations were sent to 549 349 potential participants, of which 16 389 people (3.0 percent) thought they would be eligible and completed the online screening questions. Of the respondents who completed all screening questions, 10 555 (64.4 percent) did not meet the qualifying criteria and were excluded from the remainder of the survey. Of the 10 555 respondents screened out, 8206 did not fall into one of the eight respondent categories. The remainder were excluded for reasons such as not being able to identify the NH's name, not residing in the United States, failing data fraud checks, or failing to complete the screening survey. Only 145 were excluded because they were “not familiar at all” with the quality of care in the relevant NH.

A total of 4536 (27.7 percent) of the 16 389 who accessed the screening survey met the requisite qualifying criteria and completed the full survey, which took respondents on average 11 minutes. Of these complete survey responders, 200 were eliminated because of survey wording changes during the slow rollout and 26 were missing group classifications, leaving a total sample size of 4310. In order to increase the chances that the time-quality trade-offs presented to respondents would be meaningful, 189 respondents who did not walk, drive, or use public transit or who travelled more than 60 minutes to the relevant NH were eliminated from the current analytic sample. Consequently, the analytical sample consisted of 4147 respondents. Table 2 shows the distribution of respondents across the eight respondent groups.

We initially randomized additional travel time bids in the CVM elicitation over four values (10, 20, 30, and 40 minutes). Review of the acceptance rates of bids for the first 388 respondents (9 percent) raised concerns about too low an acceptance rate (would be willing

**TABLE 1** Quality indices assessed

Quality Index (QI)	QI at current nursing home	QI at alternative nursing home	Offered absolute change in QI	SD of QI (national data)	Offered QI Change in SDs
Pressure ulcers (percent with pressure ulcer)	30	20	10	3.89	2.57
Activities of daily living 1 (percent with limitations)	20	10	10	7.03	1.42
Activities of daily living 2 (percent with limitations)	20	5	15	7.03	2.13
Pain (percent with moderate to severe pain)	10	5	5	5.45 (LS) 10.27 (SS)	0.92 (LS) 0.49 (SS)
Inspections (number of citations)	8	4	4	5.5 <sup>a</sup>	0.73
Rehospitalizations (percent rehospitalized)	25	15	10	5.95	1.68
Discharge to community (percent sent home)	55	65	10	10.55	0.95
Staffing (nursing hours per resident per day)	4	5	1	1.01	0.99

Abbreviation: SD, standard deviation.

<sup>a</sup>Mean of within-state standard deviations.

**TABLE 2** General summary statistics for model

General covariates	Mean (SD)
Physical condition [0-4 scale, where 0 is "Able to perform all daily activities without any help from another person" and 4 is "Unable to do any daily activities without assistance"]	2.1 (1.2)
Mental condition [where 0 is "No mental disabilities that affect ability to perform daily activities" to 4, where 4 is "Serious mental disability that requires assistance for all daily activities"]	1.6 (1.3)
Time in nursing home [y]	1.8 (2.9)
Actual travel time [h]	0.38 (0.30)
Respondent status	Percent of respondents
Currently residing, short-term	3.2
Currently has a family member, short-term	14.6
Currently residing, long-term	1.4
Currently has a family member residing, long-term	65.3
Was resident in last 6 mo, short-term	4.7
Had family member in short-term in last 6 mo	5.9
Discharged resident in last 6 mo, expected long-term	1.1
Had family member in long-term in last 6 mo	4.6
Scenario specific covariates	Percent
Viewed no quality index scenarios	52
Viewed one quality index scenario	16
Viewed both quality index scenarios	32
Least viewed quality index scenario: Pressure Ulcers	28
Most viewed quality index scenario: Inspections	49
Experienced neither quality index	43
Experience one quality index	45
Experienced both quality indices	12
Least experienced quality index: Discharge	1
Most experienced quality index: Activities of daily living	17

to travel the additional time for the higher quality) at 10 minutes and too high an acceptance rate at 40 minutes. Therefore, we added an additional low bid (5 minutes) and two additional high bids (50 and 60 minutes). Consequently, the remaining 3922 (91 percent) of the qualified respondents completed the survey that randomized over seven bids in the time-quality trade-off elicitations.

## 2.2 | Elicitation-related questions

Each of the eight possible elicitations had three primary components: first, a description of the quality index and an opportunity to read a related scenario; second, the actual elicitation with randomized bid of additional travel time; and third, a follow-up question to respondents accepting the bid about their certainty of acceptance. To increase sample sizes above the recommended minimum of 500 per dichotomous choice elicitation, each respondent was given two elicitations—the order of these two elicitations was randomly varied across respondents and the subsequent analysis failed to show differences due to the order of the elicitation. The following

shows the basic elicitation structure for one quality index (pressure ulcers) and one respondent group (currently has a family member residing, long-term).

The elicitation begins with a description of the quality index and an opportunity to read a scenario about it:

Think about the following problem that some nursing home patients may experience: pressure sores.

Pressure sores, sometimes also called pressure ulcers, are skin wounds that people get when they sit or lie down in the same position for a long time.

They can be very painful

They can take a long time to heal

They can cause other complications such as skin and bone infections

There are several things that nursing homes can do that may help to prevent or treat pressure sores.

Nursing homes can:

Change the position of the resident often

Make sure residents get good nutrition,

Use special mattresses to reduce pressure on the skin.

Some residents may get pressure sores even when the nursing home provides good preventive care.

Would you like to read a short example about pressure sores? [Yes, No]

Did you or your family member ever have a pressure sore or a pressure ulcer? [Yes, No]

The elicitation then describes the situation in the current nursing home and an alternative NH that is farther away:

Imagine that the "CURRENT" nursing home, where your family member is staying now, has a level of quality such that 10% of residents have pressure sores. Imagine also that it takes you 30 minutes to travel to the home every time you visit this family member and another 30 minutes to get back. Assume that you visit once per week.

In an "ALTERNATIVE" nursing home, which is farther from your home, only 5% of residents have pressure sores.

The "ALTERNATIVE" nursing home is identical in every other way to the "CURRENT" nursing home. Your family member can move to it from the "CURRENT" nursing home at no extra cost.

The elicitation then describes the choice decision where  $X$  is the randomly assigned bid time (first bracket below) and  $30 + X$  is the total minutes to the alternative nursing home (second bracket). Each elicitation presents a diagram showing lines labeled with travel times between the respondent's home and the two nursing homes.

Would you want him or her to move to this "ALTERNATIVE" nursing home that improves his or her pressure sores quality by 5%, but increases your travel time by  $[X]$  minutes from 30 minutes to  $[30 + X]$  minutes in each direction? (Please take a look at the picture below). In other words, you would be trading longer travel time for better quality.

Before asking whether the respondent would be willing to accept the additional travel time cost to gain the better quality, the respondent is reminded about the additional travel time.

Remember that you would have a travel time of  $[30 + X]$  minutes in each direction each week when you visit your family member.

Would you want your family member to move to the "ALTERNATIVE" nursing home? [Yes, No, Don't know]

The elicitation finishes by asking those who chose the alternative nursing home about the certainty of their choices.

How certain are you about your choice between the "CURRENT" and the "ALTERNATIVE" nursing homes? [Not certain at all, Somewhat certain, Pretty certain, Very certain, Extremely certain]

### 2.3 | Estimation

We estimate WTT with a model that allows us to infer mean WTT for improvements in each quality index. It also allows us to assess the effect of resident and respondent characteristics, as well as behavioral factors (such as experience with the condition or sufficient interest to read a scenario) that might affect responses.<sup>16</sup> The model also allows consideration of the feasibility of conditioning report card weights on individual characteristics. We then demonstrate how the WTT estimates can be used to construct report card weights.

A major concern in using CVM is that the hypothetical nature of the choice leads some respondents to accept bids when, if faced with actual choices, they would reject them. To guard against this sort of bias, researchers usually include a follow-up question asking about how certain respondents are of their acceptance of the bid.<sup>17</sup> "Don't know" responses and acceptances without a high level of certainty are converted to rejections. Especially with respect to private goods (like NH care in our case), where comparisons can be made between the stated CVM preferences and observed market behavior, this procedure appears to eliminate bias resulting from the hypothetical nature of the choice.<sup>18</sup> Following this approach, we estimate mean WTT after recoding acceptances as rejections if respondents were not "very" or "extremely" certain of their acceptances. We also conducted sensitivity analysis in which we did not recode acceptances for certainty.

We consider three sets of covariates that may affect WTT. First, the characteristics of the resident may be associated with the respondent's valuation of quality indices changes. In this category, we include physical condition (0-4, with 0 "Able to perform all daily activities without any help from another person" and 4 "Unable to do any daily activities without assistance"), mental condition (0-4, with 0 "No mental disabilities that affect ability to perform daily activities" and 4 "Serious mental disability that requires assistance for all daily activities"), and the number of years the resident has been (or was) in

**TABLE 3** Logit models for probabilities of bid acceptance: coefficients and (Standard errors)

	Pressure ulcers	ADL <sup>a</sup>	ADL <sup>a</sup>	Pain	Inspect	Rehospital	Discharge	Staff
Ln of minutes of added time	-0.34 <sup>b</sup> (0.094)	-0.19 <sup>b</sup> (0.10)	-0.60 <sup>b</sup> (0.098)	-0.13 (0.091)	-0.45 <sup>b</sup> (0.092)	-0.20 <sup>b</sup> (0.092)	-0.35 <sup>b</sup> (0.093)	-0.50 <sup>b</sup> (0.098)
Physical condition (0-4; 4 worst)	-0.12 (0.63)	-0.14 <sup>b</sup> (0.067)	-0.060 (0.066)	-0.082 (0.063)	0.0047 (0.059)	-0.040 (0.060)	-0.039 (0.063)	-0.15 <sup>b</sup> (0.066)
Mental condition (0-4; 4 worst)	-0.037 (0.059)	-0.053 (0.062)	-0.18 <sup>b</sup> (0.062)	-0.021 (0.057)	-0.11 (0.057)	-0.023 (0.056)	-0.066 (0.060)	0.050 (0.060)
Time in nursing home (y)	-0.028 (0.026)	-0.0074 (0.031)	-0.053 <sup>b</sup> (0.027)	0.022 (0.031)	0.025 (0.021)	-0.028 (0.029)	0.053 <sup>b</sup> (0.025)	-0.022 (0.022)
G1: residing, short-term	0.53 (0.35)	1.7 <sup>b</sup> (0.41)	0.63 (0.37)	1.4 <sup>b</sup> (0.39)	0.32 (0.38)	1.0 <sup>b</sup> (0.49)	0.31 (0.57)	1.8 <sup>b</sup> (0.42)
G2: family, short-term	0.035 (0.20)	-0.056 (0.22)	-0.14 (0.22)	0.13 (0.20)	0.15 (0.20)	0.017 (0.20)	-0.017 (0.20)	-0.17 (0.22)
G3: residing, long-term	-0.89 (0.82)	0.051 (0.61)	1.2 <sup>b</sup> (0.58)	0.39 (0.62)	-0.45 (0.47)	0.27 (0.52)	0.37 (0.65)	-0.79 (0.68)
G5: was resident, short-term	0.38 (0.32)	0.35 (0.32)	0.35 (0.32)	0.34 (0.34)	0.64 <sup>b</sup> (0.31)	0.77 <sup>b</sup> (0.30)	-0.43 (0.50)	0.62 <sup>b</sup> (0.30)
G6: family, former short-term	0.37 (0.31)	-0.16 (0.32)	0.38 (0.31)	0.87 <sup>b</sup> (0.31)	0.099 (0.29)	0.54 (0.28)	0.29 (0.31)	0.29 (0.29)
G7: was resident, expected long-term	-0.40 (1.2)	-0.97 (1.1)	NC	-1.1 (1.1)	-0.26 (0.92)	0.81 (0.93)	0.19 (0.82)	-0.33 (0.84)
G8: family, former long-term	0.38 (0.34)	0.14 (0.34)	0.015 (0.31)	0.42 (0.32)	0.59 (0.31)	0.39 (0.33)	-0.18 (0.32)	0.92 <sup>b</sup> (0.40)
Experience with quality index (1 yes, 0 no)	0.61 <sup>b</sup> (0.15)	0.36 <sup>b</sup> (0.16)	0.32 <sup>b</sup> (0.16)	0.45 <sup>b</sup> (0.15)	NA	0.28 <sup>b</sup> (0.13)	1.2 <sup>b</sup> (0.50)	NA
Viewed Scenario (1 yes, 0 no)	0.77 <sup>b</sup> (0.15)	0.62 <sup>b</sup> (0.15)	0.72 <sup>b</sup> (0.15)	0.83 <sup>b</sup> (0.14)	0.64 <sup>b</sup> (0.13)	0.31 <sup>b</sup> (0.14)	0.88 <sup>b</sup> (0.14)	0.60 <sup>b</sup> (0.14)
Actual travel time (min)	0.057 <sup>b</sup> (0.025)	0.079 <sup>b</sup> (0.030)	0.56 <sup>b</sup> (0.23)	0.031 (0.025)	0.070 (0.024)	0.033 (0.026)	0.016 (0.023)	0.056 <sup>b</sup> (0.027)
First bid (min)	0.0022 (0.0042)	0.00081 (0.0046)	0.0074 (0.0044)	0.001 (0.0041)	0.0082 <sup>b</sup> (0.0040)	-0.0025 (0.0041)	0.0045 (0.0043)	-0.000 (0.0043)
Constant	0.25 (0.31)	-0.55 (0.34)	0.75 <sup>b</sup> (0.34)	-0.91 <sup>b</sup> (0.33)	0.67 <sup>b</sup> (0.31)	-0.0080 (.31)	-0.22 (0.32)	0.77 <sup>b</sup> (0.32)
Number of observations	1032	1027	1028	1037	1036	1032	1037	1034
Model chi-squared	87.9 <sup>b</sup>	78.8 <sup>b</sup>	102.9 <sup>b</sup>	88.4 <sup>b</sup>	66.6 <sup>b</sup>	45.6 <sup>b</sup>	102.4 <sup>b</sup>	108.0 <sup>b</sup>
Coefficient of discrimination	0.09	0.08	0.10	0.09	0.06	0.04	0.10	0.10
Mean WTT (SE)	20.8 (1.7)	11.5 (3.5)	26.5 (1.1)	9.3 (5.0)	30.5 (1.7)	14.6 (4.5)	19.4 (1.2)	25.6 (0.5)

Abbreviations: NA, Not applicable (experience question not asked); NC, No cases.

<sup>a</sup>Activities of daily living.

<sup>b</sup>Statistically significant at the 5% level (bid coefficient one-sided Z test, others two-sided Z tests).



**TABLE 4** Conditional mean of willingness-to-trade travel times (95% intervals) in minutes

	Pressure ulcers	ADL <sup>a</sup>	ADL <sup>a</sup>	Pain	Inspect	Rehospital	Discharge	Staff
G1: residing, short-term N = 138	27.6 (26.8, 28.5)	14.1 (13.5, 14.6)	41.1 (38.0, 44.1)	10.1 (9.6, 10.5)	36.1 (34.9, 37.3)	16.9 (16.6, 17.2)	25.0 (23.1, 26.9)	40.1 (38.7, 41.6)
G2: family, short-term N = 619	21.3 (20.5, 22.1)	11.6 (11.2, 12.1)	26.1 (24.6, 27.5)	9.7 (9.4, 10.0)	31.9 (31.1, 32.7)	14.6 (14.4, 14.8)	19.0 (18.2, 19.8)	22.8 (21.9, 23.7)
G3: residing, long-term N = 61	13.8 (9.8, 17.7)	12.5 (11.4, 13.6)	45.8 (42.0, 49.5)	10.5 (9.5, 11.5)	27.7 (25.4, 29.9)	16.1 (15.5, 16.7)	25.6 (22.6, 28.5)	16.1 (13.3, 18.8)
G4: family, long-term N = 2813	19.7 (19.4, 20.1)	11.3 (11.1, 11.5)	24.7 (24.1, 25.3)	8.9 (8.8, 9.1)	29.2 (28.9, 29.6)	14.0 (13.9, 14.1)	19.2 (18.8, 19.6)	24.4 (24.0, 24.8)
G5: was resident, short-term N = 204	25.6 (24.4, 26.8)	13.9 (13.3, 14.6)	34.6 (32.2, 37.0)	10.4 (10.0, 10.9)	36.7 (35.9, 37.5)	17.6 (17.5, 17.7)	22.1 (19.7, 24.6)	36.4 (35.1, 37.6)
G6: family, former short-term N = 253	23.8 (22.8, 24.9)	10.2 (9.5, 10.8)	31.2 (29.1, 33.4)	11.2 (11.1, 11.4)	31.0 (29.8, 32.3)	16.7 (16.5, 16.9)	19.8 (18.6, 21.1)	25.9 (24.8, 27.0)
G7: was resident, expected long-term N = 22	18.1 (9.0, 27.2)	7.5 (4.8, 10.2)	NC	6.5 (4.5, 8.5)	26.8 (20.9, 32.7)	17.7 (17.5, 17.9)	25.6 (21.9, 29.3)	21.2 (16.0, 26.5)
G8: family, former long-term N = 200	22.8 (21.4, 24.2)	11.7 (11.0, 12.5)	23.3 (21.3, 25.4)	10.3 (9.9, 10.7)	34.9 (33.8, 36.1)	16.3 (16.0, 16.6)	16.6 (15.2, 18.0)	34.5 (33.1, 35.8)

Abbreviations: NC, No cases; N is number of respondents (each responded to two elicitation).

<sup>a</sup>Activities of daily living.

the NH. Second, we include indicators for respondent characteristics. Third, we control for several potential behavioral influences on the respondents. We include indicators for whether the respondent had experience with the quality index and whether the respondent chose to read the quality index scenario. Although the CVM scenarios clearly specified a hypothetical travel time between home and the current NH, respondents' perception of the time may be influenced by the actual travel time experienced so actual travel time was included as a covariate. In addition to these substantive covariates, the model includes the bid time for the first of the two presented CVM elicitation to determine whether elicitation order influenced responses.

We assume an underlying random utility model in which WTT is an exponential function of the bid time.<sup>12</sup> We use a standard logistic regression with the natural log of the bid time to estimate the mean and standard deviation of WTT using formulas derived by Buckland et al<sup>19</sup> Specifically, a probability distribution over bids for WTT implied by the logistic regression is numerically integrated to obtain a mean WTT conditional on the logit covariates. We implemented formulas provided by Buckland et al<sup>19</sup> for the variance of WTT with numerical integration. (See Appendix S1 for formulas.)

### 3 | RESULTS

Table 2 provides summary statistics for the covariates in the estimated model. Table 3 presents the results of the logit models for the eight quality indices as well as the estimated mean WTTs and their standard errors. In terms of model validity, note the following: First, except for Pain, there is a consistent statistically significant "price" response in that longer added travel times to the higher-quality alternative NH reduce the probability of choosing the alternative. Second, with respect to the scope test, there is a substantially larger mean WTT for the larger improvement in ADL—26.5 vs 11.5 minutes. That is, respondents were sensitive to the magnitude of change being offered. Third, the amount of the first bid only has a statistically significant effect for inspections, suggesting that employing two elicitation per respondent will in general not affect the results. Taken together, these results suggest the plausibility of CVM strategy of expressing bids for quality improvements in terms of added travel time.

Turning to the substantive covariates, the characteristics of residents have small and generally not statistically significant effects on the probability of accepting bids; better physical health reduces the probability of acceptance for ADL 1 and staffing improvements, and better mental health reduces the probability of acceptance for ADL 2 improvements. Time in nursing home reduces the acceptance rate for ADL 2 improvements but increases it for discharge improvements.

In terms of respondent characteristics, currently residing short-term (G1) has large positive statistically significant effects on the probability of accepting bids for half of the indices (ADL 1, pain, rehospitalization, and staffing) improvements and formerly residing short-term has smaller positive effects for inspection, rehospitalization, and staffing improvements; currently, residing long-term has a



large positive statistically significant effect on the probability of bid acceptance for ADL 2 improvements.

The behavioral factors operate more consistently across quality indices. Prior experience with a quality index, or choosing to read a scenario about it, has a positive statistically significant effect on probability of acceptance across the quality indices. Actual travel time has statistically significant effects for four of the quality indices (pressure ulcers, ADL 1, ADL 2, and staffing) improvements.

Sensitivity analyses found that not recoding acceptances for certainty only slightly changed the relative magnitudes of WTT so that the derived weights changed little. Mode of travel also did not affect mean WTT.

The mean WTT for each estimation sample (average of predicted values) ranges from 9.3 minutes for ADL 1 improvements to 30.5 minutes for inspection improvements. Table 4 presents mean WTT and 95 percent confidence intervals conditional on respondent status. Because being a family member with a relative in long-term care (G4) is by far the most common respondent status, its mean WTT values are close to those shown in Table 3. Relative to these values, the most striking differences are for those currently residing in short-term care, who consistently express larger mean WTTs across the quality indices. For example, whereas family members of relatives in long-term care have a mean WTT of 24.4 minutes for ADL 2, those currently in short-term care have a substantially larger value of 40.1 minutes.

### 3.1 | Demonstration of derivation of weights for composite measure calculation

To move from mean WTTs to weights reflecting preferences for use in a composite measure like the 5-Star, it is necessary to translate the CVM quality changes into the common metric of standard deviations calculated from national data. The first column of Table 5 repeats the last column of Table 1, which expresses the CVM quality changes in terms of standard deviations. The third column of Table 5 shows mean WTT per standard deviation. These values range from 8.1 minutes for a standard deviation improvement in pressure ulcer rates to 41.8 minutes for a standard deviation improvement

in inspection outcomes. Note that the mean WTT per standard deviation differs for pain depending on whether the long-term or short-term standard deviation is used—in the following illustrative calculations, we use the long-term value. Also, note that the mean WTT per standard deviation is not linear in ADL—the larger improvement has a value about 50 percent larger. As we need a single value for our illustration, we average the two ADL values.

The last column of Table 5 shows the illustrative weights based on the mean WTT per standard deviation of change for the seven quality measures standardized to sum to one. Even ignoring the very large weight for inspections, the weights show considerable variation. For example, the weight for a standard deviation improvement in pain is more than twice as large as the weight for pressure ulcers. These weights can now be applied to the quality indices of any NH to obtain its composite measure reflecting the relative importance placed on each quality index by survey respondents. This would be the alternative composite measure, which is based on patient preferences, to the 5-Star composite.

Standard CVM employs bid prices in dollars rather than minutes to obtain directly a willingness-to-pay for changes in the good being valued. Although the effect of income on willingness-to-pay in these studies is ambiguous for public goods, it should have a positive effect for private goods. Our expression of bid prices in time rather than dollars does not lead to a prediction of an effect: On the one hand, we would expect a positive income effect; on the other hand, the opportunity cost of time is higher implying an offsetting negative price effect. As a sensitivity analysis (not shown), we included income in the models shown in Table 3. In no case was the coefficient of income close to being statistically significant. The inclusion of income has negligible effects on the coefficients of the other covariates.

## 4 | DISCUSSION

This research demonstrates how CVM can be used to elicit information for developing consumer-based weights for changes in nursing home quality indices. The primary innovation is to elicit WTT in time

**TABLE 5** Willingness-to-trade (minutes) per standard deviation of quality index improvement

	SD Change of quality index in scenario	WTT	WTT/SD	Weight to be used in composite measure calculation
Pressure ulcers	2.57	20.8	8.1	0.06
ADL <sup>a</sup>	1.42	11.5	8.1	0.08
ADL <sup>a</sup>	2.13	26.5	12.4	
Pain (short-term)	0.92	9.3	10.1	NA
Pain (long-term)	0.49	9.3	19.0	0.14
Health Inspections	0.73	30.5	41.8	0.31
Rehospitalizations	1.68	14.6	8.7	0.06
Discharge	0.95	19.4	20.4	0.15
Staffing	0.99	25.6	25.9	0.19

Abbreviation: NA, Not Applicable.

<sup>a</sup>Activities of daily living.

rather than willingness-to-pay in money. Overall, the results indicate that eliciting WTT is feasible; WTT responds to bid increases as expected and the scope test shows strong responsiveness to magnitude. As only one of the valuations indicates possible influence of the first elicitation bid on acceptance of the second, it appears that respondents can be asked two elicitations with only minor risk of the first influencing the second.

The results suggest several issues that require attention, however. First, behavioral factors affect bid acceptance. Experience with the quality index appears to make it more salient to the respondent, thus increasing the probability of accepting the bid. Choosing to read a scenario about the quality index may also increase its salience, but also might indicate greater engagement with the survey content. Longer actual travel times appear to make the added time in bids appear relatively smaller, also increasing the probability of accepting a bid. As the experience and scenario effects are relatively consistent across quality indices, their impact on the derivation of relative weights is unlikely to be large. Nonetheless, until researchers develop more experience with applying CVM in this context, it is important to consider possible behavioral influences on responses. It also requires researchers to confront trade-offs between better informed respondents, and the risks of temporary increases in saliency that do not represent stable underlying preferences.

Second, the status of respondents makes a difference. Family members of residents in long-term care are most readily available for surveying. However, in general, they tend to value improvements in quality differently than current or recent residents. One might argue that we should be seeking the preferences of those currently in nursing homes. This may be feasible for short-term residents, but it is unlikely to be feasible for long-term residents without resort to costly in-person administration of surveys.

Third, the CVM requires large sample sizes—results using only the first elicitation yielding about 500 respondents per quality index are much less reliable than those based on two elicitations yielding approximately 1000 respondents per quality index. The larger sample size should be viewed as a lower bound. Consequently, applying the method to all quality indices would be costly as it would require very large sample sizes.

Fourth, currently all of the 16 included quality measures (ie, quality indices excluding inspections and staffing) are implicitly given equal weight in the assignment of the 5-Star quality component. (The remaining 8 are excluded from the 5-Star, ie, given a weight of zero.) Nursing homes earn between 20 and 100 points depending on their score on the quality index relative to the distribution of scores across nursing homes. However, the preference-based weights for quality indices are far from equal. For example, the ADL weight is more than twice that of pressure scores. This indicates a divergence between the expert opinion, which views all of the included quality indices as of equal importance, and the consumer preferences, which value some quality indices more than others.

This raises an important issue: Which set of weights should report cards use when presenting composite measures to consumers? Weights based on expert opinion or weights based on average preferences of

consumers similar to them who experienced care in the setting. Clearly, each has its advantages and disadvantages. Expert opinion is based on clinical knowledge that most consumers lack. They not only understand each of the quality indices and its shortcomings, but also the progression of disease and what patient of a particular type might face down the road. On the other hand, their weights do not reflect the patient perspective nor did they account for the variability across patients in attitudes toward disease and risk, and the constraints, financial and others, that individuals face. This research was not designed to answer the question if one set of weights is preferred to another. It was motivated by prior studies, discussed in the introduction, suggesting that consumers may find the expert opinion lacking.<sup>8-10</sup> And perhaps the best approach, in an era of transparency enabled by current information and web technology, is to offer consumers a choice, of both an “expert-based” 5-Star and a “consumer like you” 5-Star ranking.

In summary, estimating WTT may be a feasible way to create weights based on consumer preferences. Like NH care, many types of health care have multiple dimensions. Weights will be needed to create stars, letter grades, or other summaries that make the multi-dimensional information comprehensible. CVM should be explored as an approach to creating such summaries.

## ACKNOWLEDGMENT

*Joint Acknowledgment/Disclosure Statement:* This research was supported by National Institutes of Health grant R01 AG049705. We also thank Paul Nisbet and One Research for helpful advice during survey development. Debra Saliba is an employee of the Veterans Administration. The views presented here do not represent those of the Department of Veteran's Affairs.

## ORCID

David L. Weimer  <https://orcid.org/0000-0001-9237-623X>

Debra Saliba  <https://orcid.org/0000-0001-5177-6616>

Heather Ladd  <https://orcid.org/0000-0003-4427-7178>

Dana B. Mukamel  <https://orcid.org/0000-0003-4147-5785>

## REFERENCES

1. Bevan G, Evans A, Nuti S. Reputations count: why benchmarking performance is improving health care across the world. *Health Econ Policy Law* 2018;14:141-161.
2. Gormley WT, Weimer DL. *Organizational Report Cards*. Cambridge, MA: Harvard University Press; 1999.
3. Medicare.gov. Nursing Home Compare – Nursing Home Profile. <https://www.medicare.gov/nursinghomecompare/profile.html#profTab=4&ID=05A311&cmprID=05A311&loc=92617&lat=33.6387024&lng=-117.8370041&cmprDist=6.3&Distn=6.3>. Accessed January 14, 2019.
4. Werner RM, Konetzka RT, Polsky D. Changes in consumer demand following public reporting of summary quality ratings: an evaluation in nursing homes. *Health Serv Res*. 2016;51(Suppl 2):1291-1309.
5. Han X, Yaraghi N, Gopal R. Five-star ratings for sub-par service: Evidence of inflation in nursing home ratings: *Governance Studies*

- at Brookings. 2016; December: [https://www.brookings.edu/wp-content/uploads/2016/12/gs\\_20161213\\_nursing-homes.pdf](https://www.brookings.edu/wp-content/uploads/2016/12/gs_20161213_nursing-homes.pdf). Accessed September 20, 2018.
6. Centers for Medicaid and Medicare Services. Design for Nursing Home Compare Five-Star Quality Rating System: Technical Users' Guide, July 2018. 2018; <https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandCompliance/downloads/usersguide.pdf>. Accessed January 14, 2019.
  7. Williams A, Straker JK, Applebaum R. The nursing home five star rating: how does it compare to resident and family views of care? *Gerontologist*. 2016;56(2):234-242.
  8. Mukamel DB, Amin A, Weimer DL, et al. Personalizing nursing home compare and the discharge from hospitals to nursing homes. *Health Serv Res*. 2016;51(6):2076-2094.
  9. Mukamel DB, Amin A, Weimer DL, Sharit J, Ladd H, Sorkin DH. When patients customize nursing home ratings, choices and rankings differ from the government's version. *Health Aff*. 2016;35(4):714-719.
  10. Sorkin DH, Amin A, Weimer DL, Sharit J, Ladd H, Mukamel DB. Rationale and study protocol for the Nursing Home Compare Plus (NHCPlus) randomized controlled trial: a personalized decision aid for patients transitioning from the hospital to a skilled-nursing facility. *Contemp Clin Trials*. 2016;47:139-145.
  11. Arrow K, Solow R, Portney PR, Leamer EE, Radner R, Schuman H. Report of the NOAA Panel on Contingent Valuation. Vol 58 (10): Federal Register; 1993:4601-4614.
  12. Haab TC, McConnell KE. *Valuing Environmental and Natural Resources: The Economics of Nonmarket Valuation*. North Hampton, MA: Edward Elgar; 2002.
  13. Arnesen T, Trommald M. Are QALYs based on time trade-off comparable?—A systematic review of TTO methodologies. *Health Econ*. 2005;14(1):39-53.
  14. Hanemann M, Kanninen B. The statistical analysis of discrete-response CV data. In: Bateman IJ, Willis KG, eds. *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation Method in the US, EU, and Developing Countries*. New York: Oxford University Press; 1999:302-441.
  15. Berrens RP, Bohara AK, Jenkins-Smith H, Silva C, Weimer DL. The advent of internet surveys for political research: a comparison of telephone and internet samples *Polit Anal*. 2003;11(1):1-23.
  16. Weimer DL. *Behavioral Economics and Cost-Benefit Analysis*. New York, NY: Cambridge University Press; 2017.
  17. Champ PA, Bishop RC, Brown TC, McCollum DW. Using donation mechanisms to value nonuse benefits from public goods. *J Environ Econ Manage*. 1997;33(2):151-162.
  18. Blumenschein K, Blomquist GC, Johannesson M, Horn N, Freeman P. Eliciting willingness to pay without bias: evidence from a field experiment\*. *Econ J*. 2007;118(525):114-137.
  19. Buckland ST, Macmillan DC, Duff EI, Hanley N. Estimating mean willingness to pay from dichotomous choice contingent valuation studies. *J R Statist Soc: Series D (The Statistician)*. 2001;48(1):109-124.

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**How to cite this article:** Weimer DL, Saliba D, Ladd H, Shi Y, Mukamel DB. Using contingent valuation to develop consumer-based weights for health quality report cards. *Health Serv Res*. 2019;54:947-956. <https://doi.org/10.1111/1475-6773.13155>