



OPEN ACCESS

Making comparative performance information more comprehensible: an experimental evaluation of the impact of formats on consumer understanding

Olga C Damman,¹ Anco De Jong,² Judith H Hibbard,³
Danielle R M Timmermans¹

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2015-004120>).

¹Department of Public and Occupational Health and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands

²Dutch National Health Care Institute, Diemen, The Netherlands

³Department of Planning, Public Policy & Management, University of Oregon, Eugene, OR, USA

Correspondence to

Dr Olga Catherina Damman, Department of Public and Occupational Health and the EMGO+ Institute for Health and Care Research, VU University Medical Center, Van der Boerhorststraat 7, Amsterdam 1081 BT, The Netherlands; o.damman@vumc.nl

Received 26 February 2015

Revised 22 September 2015

Accepted 13 October 2015

Published Online First

5 November 2015



CrossMark

To cite: Damman OC, De Jong A, Hibbard JH, et al. *BMJ Qual Saf* 2016;**25**:860–869.

ABSTRACT

Study objectives We aimed to investigate how different presentation formats influence comprehension and use of comparative performance information (CPI) among consumers.

Methods An experimental between-subjects and within-subjects design with manipulations of CPI presentation formats. We enrolled both consumers with lower socioeconomic status (SES)/cognitive skills and consumers with higher SES/cognitive skills, recruited through an online access panel. Respondents received fictitious CPI and completed questions about interpretation and information use. Between subjects, we tested (1) displaying an overall performance score (yes/no); (2) displaying a small number of quality indicators (5 vs 9); and (3) displaying different types of evaluative symbols (star ratings, coloured dots and word icons vs numbers and bar graphs). Within subjects, we tested the effect of a reduced number of healthcare providers (5 vs 20). Data were analysed using descriptive analysis, analyses of variance and paired-sampled t tests.

Results A total of 902 (43%) respondents participated. Displaying an overall performance score and the use of coloured dots and word icons particularly enhanced consumer understanding. Importantly, respondents provided with coloured dots most often correctly selected the top three healthcare providers (84.3%), compared with word icons (76.6% correct), star ratings (70.6% correct), numbers (62.0%) and bars (54.2%) when viewing performance scores of 20 providers. Furthermore, a reduced number of healthcare providers appeared to support consumers, for example, when provided with 20 providers, 69.5%

correctly selected the top three, compared with 80.2% with five providers.

Discussion Particular presentation formats enhanced consumer understanding of CPI, most importantly the use of overall performance scores, word icons and coloured dots, and a reduced number of providers displayed. Public report efforts should use these formats to maximise impact on consumers.

BACKGROUND

Comparative performance information (CPI) about healthcare providers has been embraced in Western countries, both in efforts to empower patients and as part of competition-based healthcare reforms.^{1–4} These reports highlight variations in recommended processes of care and/or patient outcomes and experiences across healthcare providers. Especially, web-based CPI reports have become common in practice in the past 10–15 years. It has become clear, however, that CPI is only used by a small number of consumers.^{5–8} An important precondition for effective use by consumers, namely the availability of adequate information, has not always been met.^{9–10} Another reason is that many people have difficulty understanding CPI.^{2 11–13} CPI usually consists of many quality indicators, described in medical or policy jargon,^{10 14 15} and often expressed in numerical terms.¹⁵ However, many consumers lack the basic health literacy and numeracy skills needed to comprehend such scores.^{16–18} Additionally, most consumers are not motivated to rigorously study a large amount of CPI, especially

since they do not always recognise any variations between providers.¹⁹ The question of how larger groups of consumers, including those from vulnerable populations, can be supported in understanding this variability has emerged as a key issue.^{20–22}

In the past few years, several presentation formats have been tested to improve comprehensibility of CPI. For example, it has been demonstrated that explanatory frameworks for healthcare quality and word icons can help consumers when using CPI, probably because these approaches foster easy interpretations.^{23–25} Furthermore, different evaluative icons have been shown to facilitate the understanding and use of CPI.^{11 23 24 26} Another line of research, focusing on information overload in health plan choices, suggests that consumers are aided by reducing the number of providers as well as the indicators displayed.^{27–31} Although this past work provides valuable suggestions for CPI design, it remains unclear on which formats will help the largely unstudied group of consumers from vulnerable populations, such as those with lower cognitive skills. Some pilot work has been done among consumers from vulnerable groups,²⁰ but this work did not look into information comprehension. Moreover, novel presentation formats have been largely investigated separately, while combinations—as is common in real CPI—have remained relatively unexplored.

This study aimed to investigate how combinations of novel presentation formats influence CPI comprehension and use among consumers, both from vulnerable populations (ie, those with relatively low educational level, health literacy, health numeracy and patient activation) and from non-vulnerable populations. We tested the effects of (1) reducing the number of healthcare providers and (2) reducing the number of quality indicators as ways to reduce cognitive effort for consumers. In addition, we assessed the effects of (3) displaying an overall performance score and (4) evaluative symbols as ways to foster easy interpretations. The effects were assessed on different outcomes related to information comprehension and use, most importantly the correct identification of the top three performers of healthcare providers.

METHODS

Study design

This study used an experimental between-subjects and within-subjects design, in which presentation formats of CPI as provided on the website 'kiesBeter' were manipulated. This website is the Dutch national government-run website for quality of care ratings. Within subjects, the effect of a small number of providers (five nursing homes) versus a larger number of providers (20 nursing homes) was tested. Between subjects, we tested the following manipulations: the display of an overall performance score (yes vs no); the display of a small number of quality indicators

(5 vs 9); and the display of different types of evaluative symbols (star ratings, coloured dots, word icons, vs numbers and bar graphs), resulting in a 2×2×5 design.

Respondents

Respondents were recruited through an online access panel (FlyCatcher Internet Research; 20 000 panel members in total, ISO 20252- and ISO 26362-certified). Members of this panel are members of the general public who have signed up to participate in various types of surveys, to be rewarded with 'monetary points'. Panel members received an email in which they were invited to participate. A total of 2124 panel members, representative of the Dutch population regarding age, gender and the geographic area in which they lived were approached. People with lower educational level (ie, no education or only primary education) were oversampled (ie, approximately half of the sample had a low educational level), to enable us to assess the effects of our manipulations in respondents with relatively low socioeconomic status (SES) and low cognitive skills. In total, 902 panel members were included. This sample size was large enough to assess small to medium main effects and two-way and three-way interaction effects (ie, effect size of 0.10–0.25) with a statistical power of 0.80. Selected panel members were randomly assigned to one of the 20 experimental conditions, using a stratified randomisation process to ensure even distributions of gender and educational level over these conditions. Participants received a small monetary token of €1.67.

Materials

Respondents saw fictitious but realistic CPI concerning nursing homes with one absolute best provider, that is, one provider that was dominant on all indicators. We used existing CPI derived from measurements with the Dutch Consumer Quality Index (CQI) Long Term Care.³² The CQI scores varied on a scale from 1 to 4, where 1 indicated non-optimal care and 4 indicated optimal care according to patients. Based on the priority that patients gave to quality indicators in previous research,^{32 33} we selected 9 from the 14 available indicators. These indicators were safety of care, conduct of professionals, mental well-being, independency, privacy, cleaning service, meals, availability of personnel and competence of personnel. We selected nursing homes by rank ordering the 752 nursing homes with complete data, and a subsequent selection of nursing homes after every 39 cases. We ensured a selection of providers that varied on the quality of provided care. The actual scores of several nursing homes were slightly adjusted in order to derive the top three of the best-performing and the worst-performing providers. One nursing home outperformed the other nursing homes on all quality

indicators. The two other nursing homes considered second and third best had more 'better than average' scores than the other nursing homes listed. The names of nursing homes were replaced by fictitious names and the order in which the nursing homes and indicators appeared was randomly chosen. The names of the 20 nursing homes in the first set were not included in the names in the second set of five nursing homes. For each condition, a screenshot was designed in the style of the government-run *kiesBeter* website. Respondents were able to access explanations of the indicators through mouse-overs. Online supplementary appendix 1 provides examples of those screenshots.

Procedure

Respondents were provided with two screenshots: one with 20 nursing homes (the 'realistic version') and one with five nursing homes (the 'reduced version') and a sequence of questions followed. We asked respondents to imagine having to choose a nursing home for themselves or for their parents/grandparents. We instructed them to take their time in viewing information about quality of care in the different nursing homes. All participants first saw the information about 20 nursing homes, as this reflected a realistic number of providers as currently presented on Dutch websites. They were then provided with questions that assessed their comprehension and hypothetical choice. Next, respondents saw a screenshot with information about five nursing homes, and they were again asked to fill in answers to several questions. Respondents were not able to return to the page displaying CPI after they had answered our questions. Finally, respondents filled in answers to questions about sociodemographic background and cognitive skills.

Variables

Comprehension of the information

Our main outcome variable was the selection of the top three nursing homes, assessed by the following item: 'According to you, what are the top 3 best performing nursing homes?' It should be noted that what is seen as 'best' for the second and third nursing homes was no more than the sum of scores on all quality indicators. So the importance consumers might attach to the different indicators was not taken into account. Other items measuring comprehension were, 'According to you, which nursing home performs best?' (this variable had one absolute best provider) and 'According to you, which nursing home performs worst?' These three items all had a multiple choice response scale with all nursing homes from the screenshot listed; respondents could select the nursing home(s) from this scale. These questions were based on questions used in previous research testing CPI presentation.^{11 34} In addition, we also measured

respondents' more verbatim comprehension of the information displaying 20 nursing homes. These measures differed for the different experimental conditions and included questions such as 'What is the score of "Zeezicht" on the item "independency"?' and 'How does "Oosterstraat" score on the item "professional personnel" compared with "De Zonnewijzer"?' Using these questions, we assessed how consumers comprehended individual quality indicators, the overall performance score, relative performance of providers, and, if provided, specific numerical information. Three response options were developed so that items were formulated with a three-point multiple choice response scale. For example, for the item 'What is the score of "Zeezicht" on the item "independency"?' in the presentation format of numbers, the three response options were 3.96, 3.43 and 3.42. For this item in the star ratings format, the options were 'better than average', 'average', and 'worse than average'.

Choice of nursing home

We formulated one question that captured hypothetical choice: 'If you had to choose a nursing home for yourself or for one of your parents/grandparents, which one would you choose?'

Evaluations of the information

Several questions were posed to assess respondents' evaluations of (using) the information: 'How easy or hard was it for you to make a choice between the nursing homes?' response scale from 1 (very easy) to 5 (very hard) and 'I would like to use this kind of information when choosing between nursing homes', response scale from 1 (completely disagree) to 5 (completely agree).

Sociodemographic background and cognitive skills

The survey further captured educational level, subjective health literacy,^{35 36} health numeracy³⁷ and patient activation (Patient Activation Measure (PAM)).^{38 39}

Statistical analyses

Descriptive analysis was conducted to assess how the information was comprehended overall and used. Subsequently, we employed analyses of variance (ANOVA) to analyse the effects of our between-subjects manipulations. We examined the main effects and two-way and three-way interactions of the three manipulations, as well as the two-way interactions between the manipulations and educational level, subjective health literacy, health numeracy and patient activation. We used paired-sampled *t* tests to assess the effects of our within-subjects manipulation of the number of nursing homes. All analyses were performed using IBM SPSS Statistics V.20.0 with a significance level of 0.05.

RESULTS

A total of 902 respondents (response of 43%) completed our questionnaire. Table 1 shows their background characteristics. Almost half of our respondents had low educational levels. Of the respondents, 15% had inadequate subjective health literacy and 61% answered one or more of the health numeracy questions incorrectly. Of our respondents, 39% were in the two lowest PAM levels. Online supplementary appendix 2 shows the numbers of participants in each cell of the 2×2×5 design.

Table 2 shows the percentages of correct answers to our questions about CPI comprehension and use, both for the realistic screenshot and the reduced

screenshot. Overall, respondents had most difficulty answering the question ‘How does “Oosterstraat” score on the item “professional personnel” compared with “De Zonnewijzer”?’ (percentage correct was 76.6% (question only asked for the realistic version)). Almost 70% correctly selected the top three nursing homes when provided with 20 nursing homes. Notably, the percentage of respondents correctly selecting the single best nursing home (86%) differed from the percentage that correctly selected the number 1 as the top performer in the complete top three (88%) and from the percentage that chose the best nursing home (83%).

Figures 1 and 2 present the percentages of respondents correctly selecting the top three nursing homes for the different between-subjects manipulations, after viewing the realistic screenshot (figure 1) and after viewing the reduced screenshot (figure 2). Overall, more respondents correctly selected the top three nursing homes when provided with the reduced screenshot (figure 2) than when provided with the realistic screenshot (figure 1).

The results of the ANOVAs for all outcome variables in the realistic version are presented in online supplementary appendix 3. Table 3 presents the main findings from these ANOVAs. The results are described below.

Reducing cognitive effort

For the selection of the top three nursing homes, a significant difference between the realistic version and the reduced version was found ($\chi^2=90.39$; $p<0.001$). When provided with the realistic version, 69.5% of respondents correctly selected the top three, whereas with the reduced version, 80.2% did so. We also found significant differences in the same direction for the other outcome variables, that is, for the selection of the best nursing home ($t=9.12$; $p<0.001$), the selection of the worst nursing home ($t=5.15$; $p=0.001$) and the choice of nursing home ($t=8.42$; $p<0.001$). Reducing the number of quality indicators overall had less influence compared with reducing the number of providers and no significant main effects on respondents’ comprehension and use of CPI in either the realistic version (see table 3) or the reduced version.

Fostering easy interpretations

In terms of the selection of the top three nursing homes in the realistic screenshot, displaying an overall performance score had a significant main effect on respondents’ answers ($F=31.66$; $p<0.001$; see table 3). Respondents who saw an overall performance score more often selected the complete top three (79.8% correct) than respondents who were not provided with an overall score (59.5% correct; figure 1). Results in the same direction were found for the selection of the best nursing home ($F=4.10$; $p=0.043$)

Table 1 Background characteristics of respondents (N=902)

Variable	N sample	(%) Sample	Percentage of Dutch population*
Gender			
Male	468	(51.9)	49
Female	434	(48.1)	51
Age			
18–34	188	(20.8)	27
35–49	249	(27.6)	29
50–64	271	(30.0)	25
65+	194	(21.5)	19
Educational level†			
Low	438	(48.6)	–
Medium	226	(25.1)	–
High	238	(26.4)	–
Subjective health literacy‡			
Inadequate	138	(15.3)	–
Adequate	764	(84.7)	–
Numeracy§			
Zero questions correct	89	(9.9)	–
One question correct	214	(23.7)	–
Two questions correct	244	(27.1)	–
Three questions correct	355	(39.4)	–
Patient activation¶			
Level 1	128	(14.3)	–
Level 2	222	(24.7)	–
Level 3	321	(35.7)	–
Level 4	227	(25.3)	–

*Based on the 2012 data from Statistics Netherlands.

†Low educational level=no education or only primary education; average education=secondary education; high educational level=tertiary education.

‡Based on the item developed by Chew *et al.*³⁵ ‘How confident are you filling out medical forms by yourself?’ Inadequate health literacy if the answer is other than ‘extremely’ or ‘quite a bit’ on this item.

§Based on the three items developed by Schwartz *et al.*³⁷ The maximum number of questions that can be correctly answered is three.

¶Based on the mean score on 13 items from the Dutch PAM13 survey developed by Hibbard *et al.*⁴⁴ and translated into Dutch by Rademakers *et al.*³⁹ After calculating mean scores of respondents, the scores were transformed into activation scores and divided into four levels of activation, using the scorecard of Hibbard *et al.*³⁸ A higher level means higher activation.

PAM, patient activation measure.

Table 2 Percentages of correct answers to the different questions related to comprehension and use of the comparative performance information on nursing homes

Variable	Item	Twenty nursing homes (realistic version) N (%)	Five nursing homes (reduced version) N (%)
Selection of top three best-performing nursing homes	According to you, what are the top three best-performing nursing homes?		
	Number 1 correct*	792 (87.8)	877 (97.2)
	Number 2 correct*	729 (80.8)	825 (91.5)
	Number 3 correct*	670 (74.3)	742 (82.3)
	Number 1+2+3 correct*	627 (69.5)	723 (80.2)
Selection of best-performing nursing home	According to you, which nursing home performs best?	776 (86.0)	874 (96.9)
Selection of worst-performing nursing home	According to you, which nursing home performs worst?	787 (87.3)	846 (93.8)
Nursing home choice	If you had to choose a nursing home for yourself or for one of your parents/grandparents, which one would you choose?†	746 (82.7)	843 (93.5)
Comprehension of score	What is the score of 'Zeezicht' on the item 'independency'?	859 (95.2)	– (–)
Comprehension of relative performance (1)	How does 'Oosterstraat' score on the item 'professional personnel' compare with 'De Zonnewijzer'?	691 (76.6)	– (–)
Comprehension of overall score	What is the overall score of nursing home 'Aan Den Oever'?	433 (97.3)	– (–)
Comprehension of relative performance (2)	How does nursing home 'De Amberboom' score on the item 'availability of personnel' compare with the other nursing homes?	380 (83.2)	– (–)
Comprehension of numerical information	Which nursing homes score between 3.70 and 3.80 on the item 'Safety'?	759 (84.1)	– (–)
Ease of choice	How easy or hard was it for you to make a choice between the nursing homes? (percentage respondents answering easy or very easy)	511 (56.7)	681 (75.5)
Use of information in daily life	I would like to use this kind of information when choosing between nursing homes (percentage respondents (totally) agreeing)	736 (81.6)	774 (85.8)

*This indicated whether the respondent correctly identified the Number 1 nursing home as the top performer, the Number 2 as the second best performer, and the Number 3 as the third best performer.

†This indicated whether the Number 1 performer was chosen by the respondent.

and the selection of the worst nursing home ($F=30.70$; $p<0.001$). Displaying an overall performance score also significantly influenced the selection of the top three when viewing the reduced screenshot

($F=21.24$; $p<0.001$), with findings in the same direction as for the realistic screenshot, but not the other outcome variables. The use of evaluative symbols also had a significant main effect on the selection of the

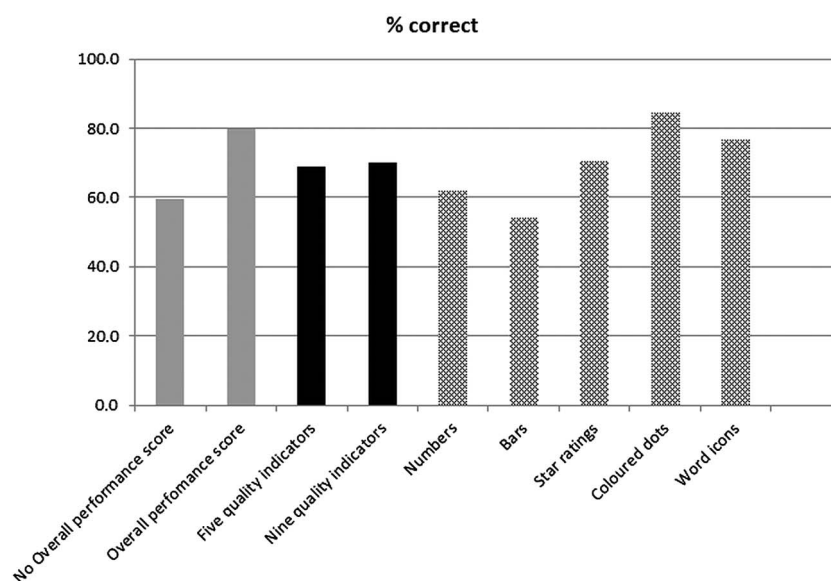


Figure 1 Percentages of respondents correctly selecting the top three best-performing nursing homes from a set of 20 nursing homes, for the different between-subjects manipulations of presentation approaches.

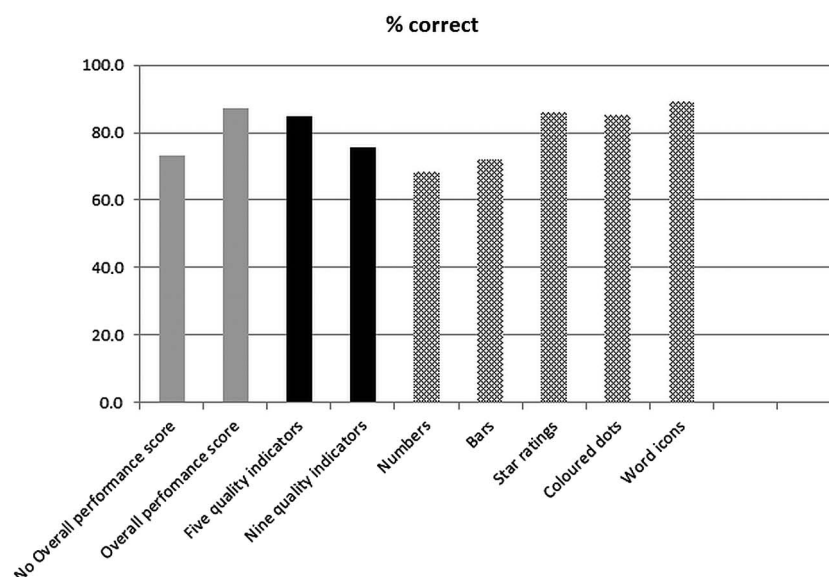


Figure 2 Percentages of respondents correctly selecting the top three of best-performing nursing homes from a set of five nursing homes, for the different between-subjects manipulations of presentation approaches.

top three with the realistic screenshot ($F=6.92$; $p<0.001$; see table 3). Respondents provided with coloured dots most often correctly selected the top three (84.3%), compared with word icons (76.6% correct), star ratings (70.6% correct), numbers (62.0%) and bars (54.2%; figure 1). Findings were in the same direction for the selection of the best

nursing home ($F=3.24$; $p=0.012$), the selection of the worst nursing home ($F=2.96$; $p=0.019$) and the nursing home choice ($F=4.16$; $p=0.002$; see table 3).

When provided with the reduced screenshot, the type of evaluative symbols significantly influenced respondents' selection of the top three nursing homes only ($F=4.19$; $p=0.002$). We also found an

Table 3 Main findings from the analyses of variance testing between-subjects manipulations of presentation formats on the different outcome variables (realistic version of 20 providers)

	Selection of top three nursing homes Effect size†	Selection of best nursing home Effect size†	Selection of worst nursing home Effect size†	Choice of nursing home Effect size†
Main effects of manipulations in presentation formats				
Display of overall performance score	0.039**	0.005*	0.038**	0.001
Small number of quality indicators	0.001	0.004	0.002	0.000
Type of evaluative symbols	0.034**	0.016*	0.015*	0.021*
Interaction effects of manipulations in presentation formats				
Display of overall performance score×small number of quality indicators	0.003	0.003	0.000	0.001
Display of overall performance score×type of evaluative symbols	0.063**	0.015*	0.010	0.007
Small number of quality indicators×type of evaluative symbols	0.018*	0.010	0.006	0.003
Display of overall performance score×small number of quality indicators×type of evaluative symbols	0.012	0.013*	0.009	0.018*
Main effects of vulnerability-related consumer characteristics				
Educational level	0.002	0.003	0.000	0.007
Health numeracy	0.066**	0.034**	0.068**	0.033**
Subjective health literacy	0.000	0.000	0.005	0.000
Patient activation	0.008	0.008	0.006	0.008

*Significant effect with significance level of 0.05.

**Significant effect with a significance level of 0.01.

†Partial eta squared, which is a measure of effect size for use in analyses of variance. The common thresholds for the magnitude of effect are 0.01=small; 0.06=medium; 0.14=large.

interaction between the display of an overall performance score and the type of evaluative symbols on selection of the top three nursing homes, both for the realistic screenshot ($F=13.17$; $p<0.001$; see [table 3](#)) and the reduced screenshot ($F=6.05$; $p<0.001$); word icons and coloured dots only supported respondents in selecting the top three when no overall performance score was displayed. This interaction was also significant and in the same direction for the selection of the best nursing home in the realistic screenshot ($F=2.87$; $p=0.022$; see [table 3](#)) and for the selection of the worst nursing home in the reduced screenshot ($F=3.27$; $p=0.011$).

Interactions between formats aimed to reduce cognitive effort and formats aimed to foster easy interpretations

For the selection of the top three in the realistic screenshot, a significant interaction was found between the number of quality indicators and the type of evaluative symbols ($F=3.64$; $p=0.006$; see [table 3](#)). Word icons and coloured dots especially aided consumers with a large number of quality indicators. The three-way interaction between the manipulations was not significant for the top three ($F=2.28$; $p=0.059$; see [table 3](#)). However, we did find significant interactions between the three between-subjects manipulations for the selection of the best nursing home ($F=2.48$; $p=0.042$) and for nursing home choice ($F=3.48$; $p=0.008$) in the realistic version (see [table 3](#)). These effects indicated that the word icons and coloured dots aided consumers when there was no overall performance score or when there was an overall performance score but a large number of quality indicators. For the reduced screenshot, we further found a significant interaction between the display of an overall score and the number of quality indicators on selection of the best nursing home ($F=4.71$; $p=0.03$), indicating that a small number of quality indicators only helped people when there was also an overall score displayed. The three-way interaction effect on the choice of nursing home was also significant with the reduced screenshot ($F=3.07$; $p=0.016$).

Consumers' vulnerability-related characteristics

Health numeracy showed a significant association with the selection of the top three nursing homes ($F=18.25$; $p<0.001$) as well as with the selection of the best nursing home ($F=9.06$; $p<0.001$), the selection of the worst nursing home ($F=19.05$; $p<0.001$) and the choice of nursing home ($F=8.73$; $p<0.001$) in the realistic screenshots (see [table 3](#)). Respondents with higher health numeracy more often adequately comprehended and used the provided CPI. We also found a significant interaction between displaying an overall performance score and health numeracy ($F=3.50$; $p=0.015$) for the selection of the worst nursing home in the realistic screenshot. The overall

performance score helped people with lower health numeracy more in selecting the worst nursing home than people with higher health numeracy. Health numeracy was also significantly related to respondents' answers when provided with the reduced screenshot, namely regarding the selection of the top three of best nursing homes ($F=3.49$; $p<0.001$), the selection of the best nursing home ($F=4.89$; $p=0.002$), the selection of the worst nursing home ($F=2.81$; $p=0.039$) and the nursing home choice ($F=2.68$; $p=0.046$).

In addition, patient activation was significantly related to respondents' selection of the top three with the reduced screenshot ($F=3.31$; $p=0.020$); respondents with the highest PAM level less often selected the top three (namely 75.5%) compared with the other three levels (level 1, 82.5%; level 2, 82.0%; level 3, 82.9%). A significant association in the same direction was found for the selection of the best nursing home ($F=5.60$; $p=0.001$) and the choice of a nursing home ($F=2.64$; $p=0.049$). A closer inspection revealed that those respondents with high PAM who showed fewer correct responses also relatively often had low health numeracy. For example, of those 47 respondents with the highest PAM level who did not correctly select the top three, 19% had the lowest numeracy level, whereas of the 145 respondents with the highest PAM level who did correctly select the top three, 6% had the lowest numeracy level. We further found an interaction effect between the number of quality indicators and subjective health literacy for the reduced screenshot ($F=5.77$; $p=0.017$), indicating that reducing the number of quality indicators had a greater positive effect in people with high subjective health literacy than among people with lower subjective health literacy.

DISCUSSION

This study investigated how different presentation formats—aimed at reducing cognitive effort and fostering easy interpretations—influence comprehension and the use of CPI in both consumers with lower SES/cognitive skills and consumers with higher SES/cognitive skills. We showed that especially the display of an overall performance score and the use of coloured dots and word icons helped consumers in comprehending and using CPI. Furthermore, a reduced number of healthcare providers displayed appeared to aid consumers. Overall, the influence of the presentation formats did not vary with vulnerability-related variables such as educational level or cognitive skills. However, there were two exceptions: (1) the display of an overall performance score appeared to particularly help people with lower health numeracy when they had to select the worst nursing home from 20 nursing homes listed; and (2) a reduced number of quality indicators appeared to only help people with

relatively high subjective health literacy when they had to select the worst nursing home from five nursing homes listed.

In line with previous studies on CPI,^{11 23 25 27} a substantial proportion of consumers had difficulties in comprehending and using the displayed CPI. For example, almost one-third of respondents (31%) were not able to correctly select the top three of best providers out of 20 listed nursing homes. Even when the number of nursing homes was reduced to five, about 20% of respondents were still not able to correctly do so. Of course we should keep in mind that for identifying the second and third best provider in our study, subjective weighing of indicators probably played a role and influenced consumers' answers. However, for the selection of the one best provider, for which we created an absolute best provider that was dominant on all indicators, 14% of respondents still failed to correctly identify the best provider. Previous studies investigating CPI found similar percentages of respondents 'miscomprehending' information.^{11 25 31 34} Studies into other types of numerical health information also found similar proportions of people not being able to correctly comprehend or use information.⁴⁰ In light of these studies, combined with the fact that less numerate people were more likely to miscomprehend information, our findings suggest that miscomprehension occurred due to difficulties in grasping the numbers as demonstrated in previous studies.^{23 40–42} However, it should be noted that most of our experimental materials did not display numbers, but rather visual displays of numbers. Even then it may be hard to perform the numerical tasks, such as computations, needed to adequately comprehend CPI.

Our findings support the notion that presentation formats can impact how consumers understand and use CPI.⁶ Especially the display of an overall performance score appeared to facilitate comprehension and use of information, with a relatively large number of providers displayed as well as a relatively small number. Furthermore, using coloured dots or word icons greatly contributed to better comprehension and use of CPI. Previous studies identified approaches such as explanatory frameworks and evaluative symbols as being supportive of consumers' comprehension and use of CPI.^{11 23–26} What our study adds to this literature is specific novel formats, such as coloured dots, that seem to be just as effective as word icons, as well as insight into the effects of combinations of these formats. Notably, displaying an overall performance score and using evaluative symbols seemed to compensate each other in providing meaning to CPI: coloured dots and word icons especially helped consumers when no overall performance score was displayed, and vice versa. In contrast, a small number of quality indicators sometimes helped people only when there was also an overall performance score displayed.

It is also interesting to note that reducing the number of healthcare providers seemed to have a greater effect compared with reducing the number of quality indicators displayed. For example, in selecting the best nursing home, reducing the number of providers displayed from 20 to 5 was associated with a 10% increase to 97% of people correctly doing so. Reducing the number of information consumers have to process has been shown to positively affect people's comprehension and use of information in numerous studies both inside the health domain^{27 31} and outside it.⁴³ It likely greatly affects the cognitive effort that consumers have to put in when weighing performance scores. For websites providing CPI, it can be difficult in practice to lower the number of providers displayed, as showing only a subset of providers might interfere with their task of creating full transparency in healthcare. A common way is to let consumers search for providers in their geographical area and to only display the ones that fall within a self-chosen distance. However, this can still leave consumers with a relatively high number of providers to compare. Showing only providers with the best performance on key indicators might be an additional option, at least as long as it is clearly communicated that such a selection has been made and that more providers can be reviewed through additional mouse clicks.

Our findings should be interpreted in light of several limitations. First, as is common with controlled experiments, this study used fictitious stimuli and hypothetical choices of consumers. Consumers' responses might not reflect their responses in real decision situations in which real CPI is used. However, the materials used were realistic and combined several different presentation formats as also used on websites. Second, we focused on 'correct' responses of consumers without taking into account the differential importance that consumers might attach to different quality indicators. While this latter approach would probably reveal more about the extent to which consumers' choices can be considered 'informed', the approach we used was an efficient way to test consumers' comprehension of information. Finally, the response rate (43%) was not very high. As is common with this type of internet panel research, it might well be that those who were not interested in quality of care hastily decided not to enrol in our study. This might be problematic because one could argue that these consumers might benefit most from viewing CPI. However, PAM levels in our study population were similar to that of the general Dutch population, so it seems that we did not capture only the responses of highly motivated consumers.

Practice implications

That consumers do not adequately comprehend and use CPI is a tremendous missed opportunity. It means that informed consumer choices are not shaping

provider behaviour in the intended direction, and that many consumers are choosing less optimal providers. Furthermore, the substantial investment in effort and resources that goes into the production of CPI is not being fully used. The evidence reported in this study can help guide more effective designs of CPI.

Acknowledgements Eliane Poort designed the experimental materials. Dave Harmsen helped in designing the experimental materials and in the data collection. He also assisted in data cleaning and statistical analyses. Uriell Malanda advised on the study design and data collection.

Contributors OCD designed the study, developed the questionnaire and collected the data. She also performed the statistical analyses, interpreted the data and drafted the manuscript. ADJ contributed to the conception and design of the study, coordinated the design of the experimental materials and data collection and helped draft the manuscript. JH contributed to the conception and design of the study and to the draft versions of the manuscript. DRMT contributed to the conception and design of the study and to the draft versions of the manuscript.

Competing interests None declared.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

REFERENCES

- Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;41(1 Suppl): I30–8.
- Sinaiko AD, Eastman D, Rosenthal MB. How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Aff (Millwood)* 2012;31:602–11.
- Taylor K. Paternalism, participation and partnership—the evolution of patient centeredness in the consultation. *Patient Educ Couns* 2009;74:150–5.
- van de Bovenkamp H, Vollaard H, Trappenburg M, *et al.* Voice and choice by delegation. *J Health Polit Policy Law* 2013;38:57–87.
- Fung CH, Lim YW, Mattke S, *et al.* Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008;148:111–23.
- Hibbard JH. What can we say about the impact of public reporting? Inconsistent execution yields variable results. *Ann Intern Med* 2008;148:160–1.
- Ketelaar NA, Faber MJ, Flottorp S, *et al.* Public release of performance data in changing the behaviour of healthcare consumers, professionals or organisations. *Cochrane Database Syst Rev* 2011;(11):CD004538.
- Teleki S, Shannon M. In California, quality reporting at the state level is at a crossroads after hospital group pulls out. *Health Aff (Millwood)* 2012;31:642–6.
- Austin PC, Anderson GM. Optimal statistical decisions for hospital report cards. *Med Decis Making* 2005;25:11–19.
- Hussey PS, Luft HS, McNamara P. Public Reporting of Provider Performance at a Crossroads in the United States: Summary of Current Barriers and Recommendations on How to Move Forward. *Med Care Res Rev* 2014;71(5 Suppl):S5–16S.
- Damman OC, Hendriks M, Rademakers J, *et al.* Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. *Health Expect* 2012;15:197–211.
- Fasolo B, Reutskaja E, Dixon A, *et al.* Helping patients choose: how to improve the design of comparative scorecards of hospital quality. *Patient Educ Couns* 2010;78:344–9.
- Kling JR, Mullainathan S, Shafir E, *et al.* Comparison friction: experimental evidence from Medicare drug plans. *Q J Econ* 2012;127:199–235.
- Schlesinger M, Kanouse DE, Martino SC, *et al.* Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. *Med Care Res Rev* 2014;71(5 Suppl):38S–64S.
- Damman OC, van den Hengel YK, van Loon AJ, *et al.* An international comparison of web-based reporting about health care quality: content analysis. *J Med Internet Res* 2010;12:e8.
- Berkman ND, Sheridan SL, Donahue KE, *et al.* Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 2011;155:97–107.
- Gaissmaier W, Gigerenzer G. Statistical illiteracy undermines informed shared decision making. *Z Evid Fortbild Qual Gesundheitswes* 2008;102:411–13.
- Paasche-Orlow MK, Parker RM, Gazmararian JA, *et al.* The prevalence of limited health literacy. *J Gen Intern Med* 2005;20:175–84.
- Rothberg MB, Morsi E, Benjamin EM, *et al.* Choosing the best hospital: the limitations of public quality reporting. *Health Aff (Millwood)* 2008;27:1680–7.
- Elbel B, Gillespie C, Raven MC. Presenting quality data to vulnerable groups: charts, summaries or behavioral economic nudges? *J Health Serv Res Policy* 2014;19:161–8.
- Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. *J Gen Intern Med* 2012;27:520–6.
- Raven MC, Gillespie CC, DiBennardo R, *et al.* Vulnerable patients' perceptions of health care quality and quality data. *Med Decis Making* 2012;32:311–26.
- Greene J, Peters E, Mertz CK, *et al.* Comprehension and choice of a consumer directed health plan: an experimental study. *Am J Manag Care* 2008;14:369–76.
- Peters E, Dieckmann NF, Västfjäll D, *et al.* Bringing meaning to numbers: the impact of evaluative categories on decisions. *J Exp Psychol Appl* 2009;15:213–27.
- Hibbard JH, Greene J, Daniel D. What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Med Care Res Rev* 2010;67:275–93.
- Hibbard JH, Slovic P, Peters E, *et al.* Strategies for reporting health plan performance information to consumers: evidence from controlled studies. *Health Serv Res* 2002;37:291–313.
- Zwijnenberg NC, Hendriks M, Damman OC, *et al.* Understanding and using comparative healthcare information; the effect of the amount of information and consumer characteristics and skills. *BMC Med Inform Decis Mak* 2012;12:101.
- Barnes AJ, Hanoch Y, Rice T. Determinants of coverage decisions in health insurance marketplaces: consumers' decision-making abilities and the amount of information in their choice environment. *Health Serv Res* 2015;50:58–80.

- 29 Hanoch Y, Rice T, Cummings J, *et al.* How much choice is too much? The case of the Medicare prescription drug benefit. *Health Serv Res* 2009;44:1157–68.
- 30 Hanoch Y, Wood S, Barnes A, *et al.* Choosing the right medicare prescription drug plan: the effect of age, strategy selection, and choice set size. *Health Psychol* 2011;30:719–27.
- 31 Peters E, Dieckmann N, Dixon A, *et al.* Less is more in presenting quality information to consumers. *Med Care Res Rev* 2007;64:169–90.
- 32 Triemstra M, Winters S, Kool RB, *et al.* Measuring client experiences in long term care in the Netherlands: a pilot study with the Consumer Quality Index Longterm Care. *BMC Health Serv Res* 2010;10:95.
- 33 Wieggers TA, Stubbe JH, Triemstra M. Ontwikkeling van een CQ-Index voor verpleeg en verzorgingshuizen en thuiszorg: kwaliteit van zorg volgens bewoners, vertegenwoordigers en cliënten. Utrecht, NIVEL, 2007.
- 34 Carman KL. Improving quality information in a consumer-driven era: showing differences is crucial to informed consumer choice. *10th National CAHPS User Group Meeting*; Baltimore, MD, 2006.
- 35 Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004;36:588–94.
- 36 Chew LD, Griffin JM, Partin MR, *et al.* Validation of screening questions for limited health literacy in a large VA outpatient population. *J Gen Intern Med* 2008;23:561–6.
- 37 Schwartz LM, Woloshin S, Black WC, *et al.* The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 1997;127:966–72.
- 38 Hibbard JH, Mahoney ER, Stockard J, *et al.* Development and testing of a short form of the patient activation measure. *Health Serv Res* 2005;40(6 Pt 1):1918–30.
- 39 Rademakers J, Nijman J, van der Hoek L, *et al.* Measuring patient activation in The Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13). *BMC Public Health* 2012; 12:577.
- 40 Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. *Am J Public Health* 2009;99:2196–202.
- 41 Peters E, Hibbard J, Slovic B, *et al.* Numeracy skill and the communication, comprehension, and use of risk-benefit information. *Health Aff (Millwood)* 2007;26:741–8.
- 42 Peters E, Hart PS, Tusler M, *et al.* Numbers matter to informed patient choices: a randomized design across age and numeracy levels. *Med Decis Making* 2014;34:430–42.
- 43 Timmermans D. The impact of task complexity on information use in multi-attribute decision making. *J Behav Decis Making* 1993;6:95–111.
- 44 Hibbard JH, Stockard J, Mahoney ER, *et al.* Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. *Health Serv Res* 2004;39(4 Pt 1):1005–26.

Experimental format with no overall performance score, nine quality indicators, numerical information, and five nursing homes.

Healthcare providers		Safety of care	Conduct of professionals	Mental well-being	Independency	Privacy	Cleaning service	Meals	Legend 3,69 1: Never good care 4: Always good care	
Provider names	Zorginstellingen	Veiligheid	Bejegening	Mentaal welzijn	Zelfstandigheid	Privacy	Schoonmaak	Maaltijd	Beschikbaarheid personeel	professionaliteit personeel
	Verpleeghuis / woon- zorg - centrum De Zwaan	3,87	3,86	3,39	3,71	3,98	3,70	3,36	3,15	3,70
	Woonzorgcentrum De Vijver	3,85	3,42	3,33	3,09	3,86	3,53	3,30	2,83	3,59
	Verzorgingshuis De Bloemen - tuin	3,95	3,91	3,61	3,74	4,00	3,81	3,70	3,29	3,81
	Verzorgingshuis Aan Den Bosrand	3,79	3,71	3,35	3,43	3,96	3,69	2,75	3,07	3,66
	Verzorgingshuis Zuiderweg	3,78	3,40	3,19	3,53	3,86	3,30	2,88	2,52	3,34

Experimental format with overall performance score, five quality indicators, word icons, and five nursing homes.

Healthcare providers		Legenda					Competence of personnel	
Provider names		Overall score	Safety of care	Independency	Privacy	Cleaning service	Professionaliteit personeel	
Zorginstellingen								
Verpleeghuis / woon- zorgcentrum De Zwaan		9,5	Beter	Gemiddeld	Gemiddeld	Beter	Beter	
Woonzorgcentrum De Vijver		9,0	Gemiddeld	Lager	Beter	Gemiddeld	Gemiddeld	
Verzorgingshuis De Bloementuin		9,7	Beter	Beter	Beter	Beter	Beter	
Verzorgingshuis Aan Den Bosrand		9,3	Gemiddeld	Gemiddeld	Beter	Gemiddeld	Beter	
Verzorgingshuis Zuiderweg		8,9	Gemiddeld	Gemiddeld	Gemiddeld	Gemiddeld	Lager	

Experimental format with overall performance score, nine quality indicators, colored dots, and twenty nursing homes.



APPENDIX: Number of participants in each cell of the 2x2x5 design

Cell	N
1	48
2	40
3	47
4	49
5	42
6	46
7	44
8	47
9	54
10	42
11	41
12	40
13	41
14	41
15	45
16	51
17	50
18	41
19	45
20	48

APPENDIX 3: Results of ANOVAs testing between-subjects manipulations of presentation formats on the different outcome variables (realistic version of 20 providers)

	Selection of top 3 nursing homes	Selection of best nursing home	Selection of worst nursing home	Choice of nursing home
	Effect size #	Effect size #	Effect size #	Effect size #
Display of overall performance score	.039**	.005*	.038**	.001
Small number of quality indicators	.001	.004	.002	.000
Type of evaluative symbols	.034**	.016*	.015*	.021*
Display of overall performance score X Small number of quality indicators	.003	.003	.000	.001
Display of overall performance score X Type of evaluative symbols	.063**	.015*	.010	.007
Small number of quality indicators X Type of evaluative symbols	.018*	.010	.006	.003
Display of overall performance score X Small number of quality indicators X Type of evaluative symbols	.012	.013*	.009	.018*
Educational level	.002	.003	.000	.007
Health numeracy	.066**	.034**	.068**	.033**
Subjective health literacy	.000	.000	.005	.000
Patient activation	.008	.008	.006	.008
Display of overall performance score X Educational level	.007	.004	.000	.001
Display of overall performance score X Health numeracy	.001	.002	.013*	.003
Display of overall performance score X Subjective health literacy	.000	.000	.000	.005
Display of overall performance score X Patient activation	.007	.003	.001	.001
Small number of quality indicators X Educational level	.001	.003	.001	.001
Small number of quality indicators X Health numeracy	.001	.002	.005	.001
Small number of quality indicators X Subjective health literacy	.000	.001	.001	.003
Small number of quality indicators X Patient activation	.003	.002	.004	.000
Type of evaluative symbols X Educational level	.015	.011	.013	.009
Type of evaluative symbols X Health numeracy	.006	.012	.022	.012
Type of evaluative symbols X Subjective health literacy	.000	.004	.009	.000
Type of evaluative symbols X Patient activation	.022	.018	.011	.011

Partial eta squared, which is a measure of effect size for use in ANOVAs. The common thresholds for the magnitude of effect are: 0.01 = small; 0.06 = medium; 0.14 = large.

* Significant effect with significance level of .05

** Significant effect with a significance level of .01