Patient Experience Journal

Volume 6 | Issue 3 Article 14

2019

Effect of change in the CG CAHPS survey instrument recall period on patient experience scores on healthcare utilization

Kimberley Marshall-Aiyelawo PhD, USN, MSC (Ret)
Office of the Assistant Secretary of Defense for Health Affairs, Falls Church, VA,
kimberley.a.aiyelawo.civ@mail.mil

Richard Bannick PhD, USAF, MSC (Ret)
Office of the Assistant Secretary of Defense (Health Affairs), Falls Church, VA,
richard.r.bannick.civ@mail.mil

Sharon Beamer AuD
US Navy, Bureau of Medicine and Surgery, Falls Church, VA, sharon.l.beamer.civ@mail.mil

Melissa Gliner PhD
US Army, Office of the Surgeon General, Falls Church, VA, melissa.d.gliner.civ@mail.mil

Terry McDavid Ms
US Army, Office of the Surgeon General, Falls Church, VA, terry.a.mcdavid.civ@mail.mil

See next page for additional authors

Recommended Citation

Marshall-Aiyelawo, Kimberley PhD, USN, MSC (Ret); Bannick, Richard PhD, USAF, MSC (Ret); Beamer, Sharon AuD; Gliner, Melissa PhD; McDavid, Terry Ms; Muraida, Daniel PhD; Ellison, Janice USAF Nursing Services (Ret); Abiero, Beatrice PhD; Roshwalb, Alan PhD; and Andrews, Mark (2019) "Effect of change in the CG CAHPS survey instrument recall period on patient experience scores on healthcare utilization," *Patient Experience Journal*: Vol. 6 : Iss. 3 , Article 14.
DOI: 10.35680/2372-0247.1359
Available at: https://pxjournal.org/journal/vol6/iss3/14

Effect of change in the CG CAHPS survey instrument recall period on patient experience scores on healthcare utilization

Cover Page Footnote

We would like to thank Amanda Sackett and Katie Zeimer for their assistance in reviewing and providing feedback on the article. This article is associated with the Policy & Measurement lens of The Beryl Institute Experience Framework. (http://bit.ly/ExperienceFramework). You can access other resources related to this lens including additional PXJ articles here: http://bit.ly/PX_PolicyMeasure

Authors

Kimberley Marshall-Aiyelawo PhD, USN, MSC (Ret); Richard Bannick PhD, USAF, MSC (Ret); Sharon Beamer AuD; Melissa Gliner PhD; Terry McDavid Ms; Daniel Muraida PhD; Janice Ellison USAF Nursing Services (Ret); Beatrice Abiero PhD; Alan Roshwalb PhD; and Mark Andrews

Patient Experience Journal
Volume 6, Issue 3 – 2019, pp. 114-123

Research

Effect of change in the CG CAHPS survey instrument recall period on patient experience scores on healthcare utilization

Kimberley Marshall-Aiyelawo, PhD, USN, MSC (Ret), *Office of the Assistant Secretary of Defense, kimberley.a.aiyelawo.civ@mail.mil*
Richard Bannick, PhD, USAF, MSC (Ret), *Office of the Asst. Secretary of Defense, richard.r.bannick.civ@mail.mil*
Sharon Beamer, AuD, *US Navy, Bureau of Medicine and Surgery, sharon.l.beamer.civ@mail.mil*
Melissa Gliner, PhD, *US Army, Office of the Surgeon General, melissa.d.gliner.civ@mail.mil*
Terry McDavid, Ms, *US Army, Office of the Surgeon General, terry.a.mcdavid.civ@mail.mil*
Daniel Muraida, PhD, *Air Force Medical Operations Agency, daniel.j.muraida.civ@mail.mil*
Janice Ellison, USAF Nursing Services (Ret), *Air Force Medical Operations Agency, janice.m.ellison.civ@mail.mil*
Beatrice Abiero, PhD, *Ipsos US Public Affairs, beatrice.abiero@ipsos.com*
Alan Roshwalb PhD, *Ipsos US Public Affairs, alan.roshwalb@ipsos.com*
Mark Andrews, *Independent Contractor, marknyllyandrews@gmail.com*

Abstract
Standardized patient experience survey instruments play an important role in informing healthcare quality and process improvement. However, any changes in standardized instruments can impact the interpretation, trending, and analysis of patient reported data. This study investigates how the change in Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG CAHPS) survey recall period, from 12- to 6-months, can impact the accuracy and quality of patient experience data. This study used primary survey data on patient experience collected in 2016. Analyses included tests of proportion and t-tests for a comparison of: 1) experience ratings, and 2) administrative data to corroborate how accurately respondents report the number of visits received within the recall period. The findings indicated that respondents, on average, underestimated their usage of care based on a 12-month recall period, apart from those who reported just one visit. A shorter 6-month recall period resulted in higher accuracy in reporting the number of actual visits that occurred. Furthermore, experiential measures showed consistently higher scores across measures for *Provider Communications*, *Staff Communications*, *Timely Access to Care*, and *Care Coordination* for a 6-month recall period compared to a 12-month period. This study showed that it would be difficult to compare CG CAHPS Version 2.0 to Version 3.0 due to recall differences in experiential measures. Given that shorter recall periods tend to be associated with higher CG CAHPS ratings, healthcare stakeholders should consider bias introduced by changes of recall periods in survey instruments.

Keywords
CAHPS, quality care, patient experience, measurement

Introduction

The Clinician and Group Consumer Assessment of Healthcare Providers and Systems (CG CAHPS) is a standardized survey methodology that has been conducted since 2007 to collect patients' experience with primary and specialty care health care providers and staff.[1] The concept of patient experience measurement is a key focus area across healthcare,[2] particularly for improving clinical care, promoting better health, and controlling healthcare costs. In 2015, the Agency for Healthcare Research and Quality (AHRQ) implemented changes to the CG CAHPS instrument.The CG CAHPS Versions 1.0 and 2.0 include a 12-month recall period, and in 2015, the CG CAHPS Version 3.0, reduced the recall period to 6-months.[1]
This study examines how the change in recall period impacts patients' self-reported experience with care. Presumably, a shorter recall period should improve the accuracy of recall due to a reduction in the number of encounters being rated, and a focus on more recent visits. The shorter recall period also reduces the risk of extraneous variance due to intervening history thereby controlling for internal validity.[3] For example, a hurricane in the past year causing severe damage and displacement of people and providers could easily distort the recall of routine access to care. Decreased cognitive burden should also reduce respondents' reliance on cognitive shortcuts, thereby reducing the influence of more general sentiments and other potentially biased inferences and allowing for a more rigorous evaluation of the encounters that they are being asked to evaluate. Independent of accuracy, the shortened recall period may also affect changes in response patterns through a literal interpretation of scale values, where the use of the end points of the scale "always" and "never" increases due to the smaller number of encounters being evaluated.
The questions guiding this study are: (a) To what extent does the reduction in the survey recall period improve

reporting of the number of visits with a given provider? (b) Does the change have an impact on experience ratings and if so, what items are affected and in what direction?

Overview

CG CAHPS

CG CAHPS is a widely used outpatient experience survey designed to monitor quality of care within organizations and, in some cases, is required by regulators and accreditation programs. CG CAHPS was the starting point for the development of the Centers for Medicare and Medicaid Service's Accountable Care Organizations (ACO) CAHPS Survey and the CAHPS Survey for the Physician Quality Reporting System (PQRS).[1]

Supplemental item sets can be added to CG CAHPS to cover various topic areas. For example, CG CAHPS with the patient-centered medical home item set (often referred to as PCMH CAHPS) is used for PCMH certification. The CAHPS consortium implemented the change in recall period to make the CG CAHPS Survey consistent across multiple stakeholders, including ACO CAHPS and PQRS.[1] A key feature of the CG CAHPS survey is that the majority of the questions refer to a specific provider.

Empirical Differences of 6- versus 12-month Recall Period and CAHPS Scale Measures

In preparation for the launch of CG CAHPS Version 3.0, AHRQ funded a randomized study in four New England health centers[4] to investigate the implications of a shortened recall period. Other than a reduction in reported care received due to the shortened recall period, Hargraves[4] found no significant differences in recall accuracy among demographic subgroups. However, the experiment showed that the 6-month recall period led to unexplained lower self-reported mental health status. Regarding experience measures, Hargraves[4] found that the 6-month recall period yielded significantly higher scores for the provider communications composite and the overall rating of the provider— the latter measure is not bounded by the recall period.

Accuracy of Reporting Periods

Several studies have investigated the effects of recall period on the accuracy of respondents reporting biographical information. Researchers have examined and tested the differences in recall periods on various topics including: employment,[5,6] income,[7] household expenditure,[8,9] patient reported outcomes for clinical trials,[10] risk behaviors and associated factors,[11–13] recreational fishing,[14,15] crime,[16] acute gastroenteritis,[17] and program evaluation.[18] Each of these studies found that shorter recall periods improved recall accuracy.

Specific to healthcare usage, Zuvekas[19] examined self-reported emergency department and outpatient visits within the Medicare beneficiary population. The study of 1,375 beneficiaries found that respondents systematically underreported both measures. For ambulatory visits, self-reported visits averaged 2.3 per 90 days as compared to 2.7 indicated by administrative data; and the error increased with the length of recall period (t-statistic = 6.61, $p<0.001$). Kjellsson et al.,[20] found similar results when they tested several recall periods on a Swedish health survey using number of days in the hospital as the main outcome. They found that inaccurate reporting fell from 2.4% (with a 12-month recall), to 0.8% (for a 6-month recall), to 0.1% (for a 3-month recall). Other researchers have reported similar underreporting of healthcare usage with longer time frames.[21–23] Bhandari and Wagner[24] conducted a systematic review of 42 empirical studies on the effect of recall period on healthcare usage and concluded that underreporting was the most common bias and that accuracy increased most notably with a reduction from a 12-month to a 6-month recall period.

Recall Periods and Response Burden on CG CAHPS Experiential Measures

The common scale of CG CAHPS questionnaire may account for the differences in performance when fewer visits are being rated. The CG CAHPS survey was designed to capture patient experience data objectively and accurately. For this reason, the CG CAHPS questionnaire primarily uses a four-point experience scale (i.e., "Never", "Sometimes", "Usually", and "Always") to capture the frequency of a given occurrence during a recall period. This four-point scale is intended to minimize subjectivity; however, one aspect of this scale is that its distribution may heavily depend on the number of visits the patient had. Patients who had only one visit in the recall period are more likely to find the options "Always"/ "Never" relevant compared to patients with multiple visits. For example, patients that saw a provider once may not have enough reference points to contemplate whether their provider "Sometimes"/ "Usually" explained things in a way that was easy to understand as would patients with multiple visits.

The extent that respondents will engage in such a strict or "accurate deliberative process" is brought into question by cognitive scientists who have demonstrated respondents often abandon strict deliberation when faced with the cognitive burden of even simple recall tasks.[25] A reduction in recall period may have a two-fold effect. Specifically, a shorter recall should reduce the cognitive burden on respondents by limiting the length of recall and the number of visits being evaluated. However, this may lead to a more literal interpretation of survey items with a 6-month recall period as described above. The effects of shorter recall should be considered further.

This paper also explores the effect of recall period on attitudinal and experience measures. Recall can be affected by many sources of error such as omission, commission,

and the recency or salience of events being recalled; however, these factors can be positive or negative for different respondents. Recall can also be biased by what Bradburn et al[25] termed *inference*—where the respondent abandons the systematic recall-and-count process, and unlocks their choices to a wide variety of external influences. Inference, to the extent that it occurs, likely undermines the intended objective functionality of the CG-CAHPS frequency scale, making the response scale function more as a satisfaction scale. Given the lack of research on these factors, this study aims to investigate only the role of bias in recall periods.

Methodology

Data and Sample

Data for this study were collected from surveys on patient experience administered by the Defense Health Agency (DHA), the medical health service arm of the United States Department of Defense. The study was approved by an Institutional Review Board (IRB number: CDO-15-2025). The United States Department of Defense provides healthcare benefits through the Military Health System to nearly 10 million active duty personnel, military retirees, and their family members. Care is delivered directly at Military Treatment Facilities (MTFs) (consisting of over 50 hospitals and 400 clinics), or via a purchased care network of civilian providers and contracted hospitals.

The DHA implemented the Joint Outpatient Experience Survey-CAHPS (JOES-C) in May 2016 using the CG CAHPS Version 3.0 protocol, which includes the 6-month recall period. Prior to May 2016, data were collected under a different program name—TRICARE Outpatient Satisfaction Survey (TROSS), which utilized the 12-month version of CG CAHPS Version 2.0. Both programs randomly selected participants with outpatient visits from each primary MTF to complete the survey. A random sample of TROSS cases (n = 2,500) and JOES-C cases (n = 2,500) were selected for this study to provide a balanced sample across MTFs. The survey program did not allow for overlapping time periods using both versions; therefore, the data used for this analysis were from adjacent months—the last two months (March and April 2016) of the TROSS program and the first two months (May and June 2016) of the JOES-C program.

The TROSS and JOES-C programs represent patients across 135 MTFs. As expected, the TROSS and JOES-C populations are very similar. Table 1 provides the population characteristics of the two sample groups. There are some unexplained differences for race (white), marital status, and post-college degree. The t-statistics provide a reference for larger versus smaller differences in percentages that are weighted and account for the sample design and non-response adjustments.

Both programs used the same sampling criteria and survey protocols—stratified design and weights to account for survey design and different response rates. The analysis was limited to primary care patient encounters to reduce any potential effects associated with visits to specialists. For assessment of respondent accuracy, the number of outpatient visits was determined directly from individual outpatient encounter records from the MHS Medical Data Repository. In other words, this administrative data provided the actual number of clinical visits each respondent had with the provider specified in the survey.

Analysis Plan and CAHPS Measures

The primary analytic goal of this study was to evaluate the effect of the change (at a 95% level of confidence) from the 12-month recall CG CAHPS Version 2.0 to the 6-month recall period in CG CAHPS Version 3.0. This included a comparison of: 1) experience ratings, and 2) how accurately respondents report the number of visits received within the recall period (see Figure 1 for how the number of visits question appears in the survey). The CAHPS experience ratings measure include: provider communications, staff communications, timely access to care, and care coordination. Each of these questions are bounded by a recall period (i.e., "in the last 12 months", or "in the last 6 months"), and utilize a four-point frequency scale of "Never", "Sometimes", "Usually", and "Always". Figure 1 also illustrates how the experiential questions appear in the survey for one of three *Access to Care* questions. Table 2 provides the exact wording for each of the CG CAHPS questions. Accuracy was evaluated by matching the reported number of visits with the provider against the actual number of visits obtained from administrative records.

Of the questions examined here, the Visit Recall question in Table 2 explores directly the accuracy of respondents' recall. The Military Health System is a closed system in which a physician or healthcare provider only sees patients within the MHS and at MHS facilities. However, in some cases, patients can obtain care via the purchased care network. Every visit a patient has with a provider is recorded in the MHS administrative records system. This allows for respondents' responses to be compared directly with their actual number of visits, giving rates of accuracy and average error. This was not the case for patients with purchased care. Thus, this study focused on patients only receiving care at MHS facilities.

Figure 1. CG CAHPS Visit Recall and Experiential Questions

Version 2.0 – 12-month Recall Period	**Version 3.0 – 6-month Recall Period**
CG CAHPS Visit Recall Questions	
In the last 12 months, how many times did you visit this provider to get care for yourself? ☐ None → **If None, go to #26 on page 4** ☐ 1 time ☐ 2 ☐ 3 ☐ 4 ☐ 5 to 9 ☐ 10 or more times	In the last 6 months, how many times did you visit this provider to get care for yourself? ☐ None → **If None, go to #23 on page 4** ☐ 1 time ☐ 2 ☐ 3 ☐ 4 ☐ 5 to 9 ☐ 10 or more times
Example Experiential Question - Access to Care Question	
In the last 12 months, when you phoned this provider's office to get an appointment for **care you needed right away**, how often did you get an appointment as soon as you needed? 1☐ Never 2☐ Sometimes 3☐ Usually 4☐ Always	In the last 6 months, when you contacted this provider's office to get an appointment for **care you needed right away**, how often did you get an appointment as soon as you needed? 1☐ Never 2☐ Sometimes 3☐ Usually 4☐ Always

Results

More Accuracy for Shorter Recall Period and Among Fewer Reported Visits

The distribution of the self-reported number of visits with provider (Table 3) is similar between 12-month recall and a 6-month recall except for the significant difference (p<0.001) for "1 Time" reported visit. The shorter recall period has a greater percentage of a single visit. On the other hand, respondents underestimate the number of visits more with a 12-month recall than with a 6-month recall, except for those who reported just one visit. For instance, of those that indicate they had visited the provider 4 times, the mean actual visits for the 12-month recall is 5.17 visits compared to 4.28 visits for the 6-month recall survey. Those who report just one visit underestimate the number of visits more on the 6-month recall survey than the 12-month recall survey. Only respondents (n = 39) who report 10 or more visits in the 6-month recall survey overestimated the visits, with an average number of actual visits of 7.97.

Table 3 also summarizes the accuracy rate of responses for each number of visit response category. Accuracy rate is defined by percentage of people who self-report a number of visits that matches the actual number of visits in the administrative data. The accuracy rate for those that reported just one visit is the highest across all categories and those with the 12-month recall were more accurate than those with the 6-month recall period—77.67% compared to 68.25%, respectively. Overall, 41.12% of the 6-month recall period responses were accurate versus 37.73% of 12-month recall period responses, and this difference is statistically significant (p=.007). The rate of "Don't know/NA" responses were less for the 6-month recall period (5.14%) versus the 12-month (9.13%) [p=.088].

Survey respondents can choose an exact visit number in the answer choices up to the fourth visit. However, after this, visit numbers are grouped as "5 to 9" and "10 or more times." These last two response options ("5 to 9" and "10 or more") make it challenging to compare reported visit number to actual visit number. In this analysis no error was recorded if the actual number of visits was in "5 to 9" visit range and recorded by the respondent, and an error was recorded if the number of visits was not in this range. The same if the actual number of visits was 10 or more and the number of visits was 10 or more. Table 3 examines the average error between the 12-month and 6-month recall. To calculate an average error, the "5 to 9" response was recoded to be "5" when the actual number of visits was 4 or less and recoded to "9" when the actual number of visits was 10 or greater. For the "10 or more times" response, those with 10 visits were assigned if the actual number of visits was fewer than 10, and a zero error when the actual number of visits was 10 or greater. This method minimized the amount of error calculated between the reported number of visits and the actual number of visits. If we were to split the results into the full range from "none" (or no visits) to "10 or more visits", the average error for the 12-month recall is -0.689 visits while the 6-month recall has less error with an average error -0.460, a statistically different result (p-value <0.001). When looking at the range where a respondent is asked to self-report either 0, 1, 2, 3, or 4 visits, the average error increases for both the 12-month and 6-month recall periods with average errors of -0.791 and -0.574 visits for

Table 1: Population Characteristics of TROSS 12-Month and JOES-C 6-Month Recall Groups

Demographic	Category	Version 2: 12-Month Recall	Version 3: 6-Month Recall	t-Statistic
Gender	Male	46.7	50.7	-1.347
	Female	53.3	49.3	1.347
Ethnicity	Hispanic	11.9%	11.2%	0.337
Race	White	79.6%	74.3%	1.999
	Black or African American	17.2%	17.1%	0.057
	Asian	9.7%	9.0%	0.479
	Native Hawaiian or Pacific Islander	2.3%	2.0%	0.460
	American Indian or Alaskan Native	2.5%	2.7%	-0.235
Marital Status	Married	77.0%	84.9%	-2.871
Age	18 to 24	8.8%	7.3%	0.796
	25 to 34	22.7%	21.8%	0.215
	35 to 44	18.9%	22.6%	-1.568
	45 to 54	19.3%	21.5%	-0.947
	55 to 64	22.8%	19.7%	1.657
	65 to 74	4.3%	4.1%	0.322
	75 or older	3.2%	3.0%	0.427
Education	1- 8th Grade or less	0.3%	0.2%	0.221
	2- Some High School or Less	1.1%	0.6%	1.961
	3- High School Graduate or GED	13.2%	12.8%	0.155
	4- Some College or 2-year Degree	39.7%	35.0%	1.681
	5- 4-Year College Degree	19.8%	18.1%	0.759
	6- More than 4-Year College Degree	25.9%	33.3%	-2.570

Note. Race categories are not mutually exclusive.

12-month and 6-month recall periods respectively. Negative average errors indicate that respondents underestimate the number of visits they have with their providers.

Changes in Response Distributions for Experiential Questions

Change in Recall Period Affects Provider Communication Questions. CAHPS experiential questions use the scale points of "Always", "Usually", "Sometimes" and "Never". Table 4 reports percentages for "Always" and "Usually" separately, and "Sometimes" and "Never" combined for each of the CG CAHPS experiential questions. The purpose is to investigate whether 12-month recall versus 6-month recall use the scales differently and for all questions. The questions are separated into the provider communications, staff communications, timely access to care, and care coordination dimensions.

As Table 4 shows with respect to the bottom-two ratings, while all but one measure reflect numerically lower ratings for 6-month respondents, the percent responding either "Sometimes" or "Never" is statistically different for each of the Provider Communications questions— "Provider Explained Things" (p=.009), "Provider Listened Carefully" (p=.003), "Provider Showed Respect" (p=.010), and "Provider Spent Enough Time" (p=.021). None of the other questions show a statistically significant difference.

Top-box results (the percent responding "Always") see larger or equal percentages for 6-month versus 12-month recall for all questions except "Provider Listened Carefully". None of the results are statistically different between the 12-month and 6-month recall periods at the 95% level of confidence. However, the Access-to-Care questions "Appointment for Urgent Care" (p=.054) and "Medical Questions Answered the Same Day" (p=.077) showed increases in top-box scores albeit not significant. It should be noted that for respondents with just one visit, responses of "Always" or "Never" would be the most appropriate scale points in most circumstances. On the other hand, results show numerically, but not statistically, equal or smaller increases for 6-month top-box ratings for 10 of the 11 questions compared to the 12-month scores. Although only two of the 11 top-box measures show marginally statistical differences, the probability of observing ten out of 11 measures change in the same direction, at random, is only 0.006.

Table 4 also includes the middle scale point to understand if patients use this scale point differently in a 12-month recall versus a 6-month recall. The Provider Communications questions have higher percentages for

Table 2. Wording for CG CAHPS Experiential Questions by Domain

Domain	Question Wording
Provider Communications*	
Provider Explained Things	In the last 12/6 months, how often did this provider explain things in a way that was easy to understand?
Provider Listened Carefully	In the last 12/6 months, how often did this provider listen carefully to you?
Provider Showed Respect	In the last 12/6 months, how often did this provider show respect for what you had to say?
Provider Spent Enough Time	In the last 12/6 months, how often did this provider spend enough time with you?
Staff Communications	
Helpful Office Staff	In the last 12/6 months, how often were clerks and receptionists at this provider's office as helpful as you thought they should be?
Courtesy and Respect	In the last 12/6 months, how often did clerks and receptionists at this provider's office treat you with courtesy and respect?
Access to Care	
Appointment for Urgent Care	In the last 12/6 months, when you contacted this provider's office to get an appointment for care you needed right away, how often did you get an appointment as soon as you thought you needed?
Appointment for Routine Care	In the last 12/6 months, when you made an appointment for a check-up or routine care with this provider, how often did you get an appointment as soon as you thought you needed?
Medical Question Answered the Same Day	In the last 12/6 months, when you contacted this provider's office during regular office hours, how often did you get an answer to your medical question that same day?
Coordination of Care	
Provider Knows Medical History	In the last 12/6 months, how often did this provider seem to know the important information about your medical history?
Follow up on Tests	In the last 12/6 months, when this provider ordered a blood test, x-ray or other test for you, how often did someone from this provider's office follow up to give you those results?

**Note that two questions that appeared in Version 2.0 from the Provider Communications Composite are no longer part of that section in Version 3.0. Specifically, "In the last 12 months, how often did this provider give you easy to understand information about these health questions or concerns?" became a supplemental question. There was also reduced length of the core measures from 34 to 31. Detailed overview of all changes to the questionnaire have been documented elsewhere.[28]*

each question, but this pattern reverses for all other questions. There is only one question, "Appointment for Routine Care," with a significant result.

Discussion

This study examined how a 12-month recall versus a 6-month recall period affected patient experience question response patterns. The findings indicated that respondents, on average, underestimated their usage of care and that a shorter recall period results in greater accuracy relative to known administrative data. Specifically, respondents providing a 12-month recall underestimated their number of visits by an average of 0.689 visits and those with a 6-month recall underestimated by an average of 0.460 visits. This difference of 0.229 visits was statistically significant and serves as an indication of differences in how respondents report their visit experience. When the reported visit categories were closely examined, the difference between average number of actual visits and the reported number of visits was always smaller for the 6-month versus the 12-month recall for response categories: "None", "1 time", "2", "3", and "4". Interestingly, for those reporting one visit, the 12-month accuracy rate was higher than the 6-month. The better recall suggests saliency of the single visit over a longer period (e.g., an annual physical) is higher than multiple visits.

These results align with previous studies that have examined the accuracy of reported versus actual visits.[20–24] Previous recall studies indicate that reducing the recall period from 12-months to 6-months likely reduces the cognitive burden on respondents. In both theoretical[24,25] and empirical[12,19,20,24] studies, errors associated with recall can be reduced by adopting a shorter recall period.

Regarding experiential questions, differences were found for provider- and staff-related questions. We used consecutive months where month-to-month differences were likely minimal. Our comparison of results showed a

Table 3: Analysis of the Number of Reported Number of Visits versus the Accuracy of Based on the Actual Number Visits for 12-Month and 6-Month Recall

Number of Reported Visits on Survey	CAHPS Version 2.0: 12-Month Recall Period				CAHPS Version 3.0: 6-Month Recall Period				Comparison of Reported Times Visited Provider+ *(p-value)*	Comparison of Accuracy+ *(p-value)*
	Respon-dents (n)	Percent (1)	Mean Number of Actual Visits	Accuracy (%) (2)	Respon-dents (n)	Percent (3)	Mean Number of Actual Visits	Accuracy (%) (4)	(1) versus (3)	(2) versus (4)
None	77	3.4	1.57	0.00	65	2.8	1.51	0.00	0.413	--
1 Time	618	27.2	1.39	77.67	907	38.4	1.61	68.25	0.001	0.001
2	560	24.7	2.70	32.68	651	27.6	2.52	31.18	0.252	0.301
3	409	18.0	4.13	22.98	341	14.4	3.57	20.82	0.188	0.280
4	259	11.4	5.17	13.51	198	8.4	4.28	20.20	0.288	0.028
5 to 9	278	12.3	7.16	47.12	159	6.7	6.16	49.06	0.067	0.266
10 or more times	68	3.0		27.94	39	1.7	7.97	30.77	0.669	0.378
Total without DK/NA	2269		3.14	42.52	2360		3.09	43.55		0.104
Don't Know/ NA	228		2.95	9.13*	128		1.40	5.14*		0.088
Total	2497		3.61	37.73	2488		2.74	41.12		0.007

**Percent of cases with Don't Know or No Answer*

+ Unweighted test for proportion

Note. P-value not applicable for change in accuracy for "None". For the full sample the average error between reported and actual visits for 12-month recall was -0.689 and for 6-month recall was -0.460 with a p-value of 0.001 based on an unweighted t-test. The average error between reported and actual visits for four or fewer reported visits for the 12-month recall was -0.791 for 6-month recall was -0.574 with a p-value of 0.001 based on an unweighted t-test.

significant change in the "Sometimes and Never" frequency ratings for the provider communications domain, but not in the staff communications, access-to-care and coordination-of-care domains. This was consistent with Hargraves[4] who found similar results in the provider communication composite. Although the survey asks respondents to report on their experience during the survey recall period, in our sample, 69.3% of the respondents reporting one visit in the recall period responded "Usually" or "Sometimes" for at least one question. This indicates that respondents appear to generalize beyond the one visit. As discussed previously, patients who had one visit were unlikely to use the middle response categories— "Sometimes" and "Usually" and the response categories "Never" and "Always" are probably less relevant to patients the more visits they have had.

Limitations

This study is not without limitations. It is important to note other changes made to the Version 3.0 survey including: modifications in composite measures that asked about access to care and provider communication, reduction in the length of the core survey items from 34 to 31, and minor re-wording of some of the items. Although we did not test the impact of reduced length or wording, a CAHPS Consortium analyses showed that the internal consistency and site-level reliability of the access and communication composites was similar and acceptable for both Version 20 and 3.0.[26] The study findings also suggest that recall more than other survey changes likely explain differences in patient reporting.

Our analyses dealt with the patient population, but we did not and cannot assume that these effects are consistent across sub-populations, such as different age groups, or heavy versus low utilizers of the MHS. The last point also means we did not account for respondents who see multiple providers and the effect of recall period. In the MHS patients often receive primary care through a

Table 4. Comparison of 12-Month versus 6-Month

Measure	Survey	Top-Box - Always		Second-Box - Usually		Bottom-2 Box - Sometimes & Never		Base Size
		Percent	*p-value**	Percent	*p-value**	Percent	*p-value**	
Provider Communications								
Provider Explained Things	12-month	80.2	0.853	12.0	0.216	7.9	0.009	2,188
	6-month	80.2		15.4		4.4		2,301
Provider Listened	12-month	79.2	0.995	11.7	0.057	9.1	0.003	2,183
	6-month	78.7		16.2		5.1		2,298
Provider Showed Respect	12-month	83.7	0.237	8.4	0.651	7.9	0.010	2,185
	6-month	85.7		9.7		4.6		2,299
Provider Spent Enough Time	12-month	75.7	0.796	16.0	0.323	8.3	0.021	2,178
	6-month	75.7		18.8		5.6		2,290
Staff Communications								
Helpful Office Staff	12-month	58.1	0.351	29.7	0.760	12.2	0.351	2,189
	6-month	60.7		28.7		10.6		2,294
Treat with Courtesy and Respect	12-month	71.2	0.118	20.7	0.543	8.0	0.072	2,189
	6-month	76.3		19.2		4.5		1,341
Access to Care								
Appointment for Routine Care	12-month	46.3	0.212	34.7	0.048	19.0	0.501	1,578
	6-month	50.0		29.4		20.7		1,545
Appointment for Urgent Care	12-month	47.0	0.054	29.2	0.276	23.8	0.289	1,089
	6-month	52.6		26.2		21.2		1,196
Medical Question Answered the Same Day	12-month	42.7	0.077	29.7	0.279	27.6	0.348	691
	6-month	49.6		24.5		25.8		873
Coordination of Care								
Provider Knew Important Medical History	12-month	59.2	0.158	24.5	0.730	16.4	0.145	2,169
	6-month	62.9		23.2		13.9		2,288
Follow Up on Tests	12-month	55.0	0.251	15.2	0.364	29.7	0.546	1,582
	6-month	58.8		13.5		27.7		1,571

**Tests for proportion to determine statistical differences accounting for weighting and design effects*

primary care team or may see specialty care providers in the same clinic or through the purchased care network. These other issues raise the question of how patients filter and cognitively process visits with more than one provider when they are asked to recall the number of visits and experiential questions with one provider. Furthermore, results are not generalizable outside of the MHS given that the patient population tends to be younger and healthier compared to the civilian population and because the study sample is restricted to visits in a closed system that operate under the same TRICARE health plans.

Contributions and Next Steps

Study results of MHS patients are consistent with civilian studies regarding recall accuracy of healthcare usage.[24] A more thorough exploration into response differences by different demographic groups might shed additional insight into the cognitive processes of such groups. Many studies show that older respondents are more prone to underestimating usage while younger respondents are more likely to overstate events.[27] Our study findings suggest that CG CAHPS should consider how the number-of-visits question should be asked, and whether it

should continue to be used given the high recall inaccuracies found among survey participants. Consequently, comparing CG CAHPS Version 2.0 to Version 3.0 should be done cautiously.

References

1. Agency for Healthcare Research and Quality. CAHPS Clinician & Group Survey. https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html. Published October 2018. Accessed November 21, 2018.
2. Wolf JA, Niederhauser V, Marshburn D, Lavela SL. Defining Patient Experience. *Patient Experience Journal*. 2014;Volume 1(Issue 1):7-19.
3. Campbell DT, Stanley JC. *Experimental and Quasi-Experimental Designs for Research*. Vol.2. Boston,MA: Houghton Mifflin Comp; 1967.
4. Hargraves L. *Sampling and Respondent Recall: What Difference Does 6 Months Make? Introducing the New CAHPS Clinician & Group Survey 3.0*. Washington, DC: Agency for Healthcare Research and Quality; 2015:25-37.
5. Dex S, McCulloch A. The reliability of retrospective unemployment history data. *Work, Employment & Society*. 1998;12(3):497-509.
6. Song Y. Recall bias in the displaced workers survey: Are layoffs really lemons? *Labour Economics*. 2007;3(14):335-345. doi:10.1016/j.labeco.2006.05.001
7. Kapteyn A, Ypma JY. Measurement Error and Misclassification: A Comparison of Survey and Administrative Data. *Journal of Labor Economics*. 2007;25(3):513-551. doi:10.1086/513298
8. Bradburn NM. Recall Period in Consumer Expenditure Surveys Program. 2010. http://www.bls.gov/cex/methwrkshp_pap_bradburn.pdf.
9. Cantor D. Discussion of Plans for Designing the Recall Period for the Consumer Expenditure Interview Survey. *Consumer Expenditure Survey Methods Workshop, US Bureau of Census*. 2010.
10. US Food and Drug Administration, US Dept. of Health and Human Services. *Guidance for Industry Use in Medical Product Development to Support Labeling Claims Guidance for Industry*. Clinical/Medical Federal Register; 2009:1-39.
11. Arnold BF, Galiani S, Ram PK, et al. Optimal recall period for caregiver-reported illness in risk factor and intervention studies: A multicountry study. *Am J Epidemiol*. 2013;177(4):361-370. doi:10.1093/aje/kws281
12. Ekholm O, Strandberg-Larsen K, Grønbæk M. Influence of the recall period on a beverage-specific weekly drinking measure for alcohol intake. *Eur J Clin Nutr*. 2011;65(4):520-525. doi:10.1038/ejcn.2011.1
13. McElrath K, Chitwood DD, Griffin DK, Comerford M. The consistency of self-reported HIV risk behavior among injection drug users. *Am J Public Health*. 1994;84(12):1965-1970. doi:10.2105/AJPH.84.12.1965
14. Fisher WL, Grambsch AE, Morganstein DR. Length of recall period and accuracy of estimates from the national survey of fishing, hunting, and wildlife-associated recreation. *American Fisheries Society Symposium*. 1991;12:367-374.
15. Connelly NA, Brown TL. Effect of recall period on annual freshwater fishing effort estimates in New York. *Fisheries Management and Ecology*. 2011;18(1):83-87. doi:10.1111/j.1365-2400.2010.00777.x
16. G. Skogan W. Methodological Issues in the Study of Victimization. In: *From Crime Policy to Victim Policy: Reorienting the Justice System*. Hampshire, UK: Palgrave Macmillan; 1986:80-116. doi:10.1007/978-1-349-08305-3_5
17. Cantwell LB, Henao OL, Hoekstra RM, Scallan E. The effect of different recall periods on estimates of acute gastroenteritis in the United States, FoodNet Population Survey 2006–2007. *Foodborgne Pathog Dis*. 2010;7(10):1225-1228. doi:10.1089/fpd.2010.0567
18. Ellis R, Chenevert R. "I Don't Remember": Effects of Recall Period on Reported Job and Program Participation Duration. In: Washington, DC: U.S. Census Bureau; 2012.
19. Zuvekas SH. The effects of recall length and reporting aids on household reporting of health care events in the medical expenditure panel survey. *JESM*. 2011;36(4):321-343. doi:10.3233/JEM-2011-0348
20. Kjellsson G, Clarke P, Gerdtham U-G. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ*. 2014;35:34-46. doi:10.1016/j.jhealeco.2014.01.007
21. Cohen SB, Wun L-M. A comparison of household and medical provider reported health care utilization and an estimation strategy to correct for response error. *JESM*. 2005;30(2,3):115-126.
22. Seidl H, Meisinger C, Kirchberger I, Burkhardt K, Kuch B, Holle R. Validity of self-reported hospital admissions in clinical trials depends on recall period length and individual characteristics. *J Eval Clin Pract*. 2016;22(3):446-454. doi:10.1111/jep.12506
23. Schmitz MF, Russell DW, Cutrona CE. The validity of self-reports of physician use among the older population. *J Appl Gerontol*. 2002;21(2):203-219. doi:10.1177/07364802021002005
24. Bhandari A, Wagner T. Self-reported utilization of health care services: improving measurement and accuracy. *Med Care Res Rev*. 2006;63(2):217-235. doi:10.1177/1077558705285298
25. Bradburn NM, Rips LJ, Shevell SK. Answering autobiographical questions: The impact of memory and inference on surveys. *Science*. 1987;236(4798):157-161.

26. AHRQ. *Comparability of Results for the 2.0 and 3.0 Versions of the CAHPS Clinician & Group Survey*.; :1-4. https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/Comparison_of_CG-CAHPS_v3_and_v2.pdf.
27. Woodside A, Wilson E. Respondent inaccuracy. *Journal of Advertising Research*. 2002;42:7-18. doi:10.2501/JAR-42-5-7-18
28. AHRQ. *CAHPS® Clinician & Group Survey: Overview of the Questionnaires*.; 2017.