Title
An Evaluation of Consumer Analysis of Healthcare Providers and Systems Patient Satisfaction Surveys at an Academic Medical Center Comparing In-Person and Telehealth Care

Author
Gotthardt, Christine Jacqueline

Publication Date
2022

Peer reviewed|Thesis/dissertation

An Evaluation of Consumer Analysis of Healthcare Providers and Systems Patient Satisfaction Surveys at an Academic Medical Center Comparing In-Person and Telehealth Care

By

CHRISTINE JACQUELINE GOTTHARDT

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Epidemiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA, DAVIS

Approved:

____

James P. Marcin, Chair

____

Sarah C. Haynes

____

Danielle J. Harvey

Committee in Charge

2022

Acknowledgements

To my wonderful thesis committee for advising me, to Ray Troche for guiding me, and to my family and Charles for supporting me.

Abstract

Consumer Analysis of Healthcare Providers and Systems (CAHPS®) patient satisfaction surveys have been frequently used to evaluate patient satisfaction for both in-person and telehealth care. This thesis had three objectives: 1) to determine whether there are differences in patient satisfaction with telehealth visits between departments, age categories, genders, ethnicities, and primary languages; 2) to compare patient satisfaction survey scores between telehealth and in-person visits and ascertain whether these two methods of care are comparable; and 3) to evaluate different approaches of categorizing CAHPS top box scores using statistical analysis. To do this, CAHPS patient satisfaction survey scores from a large patient population across various clinical departments at an academic medical center for both in-person and telehealth visits were examined. Using binary logistic regression models, this study examined patient satisfaction with telehealth among different strata of age, race, gender, ethnicity, language, and clinical department; and compared patient satisfaction with their care provider between in-person and telehealth patients in models including these covariates. Furthermore, using both binary and ordinal logistic regression models, this study examined multiple distinct categorizations of CAHPS survey scores using statistical techniques. Findings indicate satisfaction scores following patients visiting their care provider over telehealth visits are comparable to visits provided in-person. They also indicate that different categorizations of top box scores can influence the fit of logistic regression models to the data, and that this should be considered by future researchers. Further qualitative analyses of patient comments on CAHPS surveys, as well as an investigation into patient satisfaction with their care provider and overall experience with their care by clinical department, should be considered.

Table of Contents

Section 1 - Telehealth

Introduction

The use of telehealth—the provision of health services using remote technologies[1]—dramatically increased at U.S. medical centers following the start of the Covid-19 pandemic in the spring of 2020.[2] The demand for telehealth services rapidly increased worldwide with stay-at-home orders and remained even after healthcare facilities reopened.[3] Though incidence numbers have since fallen from peak epidemic levels,[4] telehealth encounters continue to comprise a sizable proportion of all ambulatory visits, particularly for patients that are immunocompromised or cannot regularly visit their local health center, or for certain visit types such as follow-up appointments.[5] As such, it is essential to understand patient satisfaction and experience with telehealth as it is adopted into regular medical care.

Consumer Analysis of Healthcare Providers and Systems (CAHPS®) surveys from the Agency for Healthcare Research and Quality have been used since the 1990s to assess quality of care from care providers, clinics, and healthcare plans.[6] The usefulness of CAHPS survey scores in evaluating patient satisfaction with many different facets of medical care has been confirmed by various studies.[7-9] The Ambulatory Clinics version of the CAHPS survey, which is the version used in this thesis, was administered following most ambulatory visits at UC Davis Health via email, text message, or both.[10] Patients were sent two follow-ups requesting their response to the survey after initial communications. The objective of this section was to use patient satisfaction survey scores from the Ambulatory Clinics CAHPS survey during the height of the pandemic to determine the extent to which patients are or are not satisfied with telehealth, and to discover any differences in satisfaction between strata of covariates at UC Davis Health.

Methods

Section 1 – Telehealth: This section is a retrospective, observational study examining CAHPS patient satisfaction survey scores for patients that met with their care provider via telehealth within UC Davis Health. For the purposes of measuring overall patient satisfaction with telehealth, three questions were selected for analysis: the patient's *ease* of talking with their care provider, how well the *video connection* worked during the visit, and how well the *audio connection* worked during the visit. All questions were scored on a 5-point Likert scale, with 1 being "Very Poor" and 5 being "Very Good".

Section 1 only includes patients that visited their care provider via telehealth. All non-surgical clinics and departments were included for study. Visits in surgical clinics were not included as the use of telehealth for surgery patients and those within surgical departments tends to be for specific encounter types that are not necessarily consistent with the use of telehealth among ambulatory specialties.[11] Moreover, patients under the age of 18, as well as patients visiting pediatric clinics, were not included in the analysis, as the pediatric CAHPS survey is different from the adult survey, and may be filled out by a parent or guardian rather than the individual patient.[12] Smaller departments—defined as those that did not make up 1% or more of the patient population—were excluded as these small clinics represent unique specialty clinics where findings may not be generalizable or have external validity. Small, specialty clinics that made up their own departments were also excluded for the same reasons. A summary of the implementation of inclusion and exclusion criteria employed to arrive at the study population can be found in **Figure 1.**

Using Stata 17.0, logistic regression models were constructed, including age, race, gender, ethnicity, language, and department as key variables of interest. Age was categorized as per AHRQ guidelines.[13] Races, ethnicities, and languages that did not make up 5% or more of the patient population were grouped into a respective "Other" category. The inclusion of these variables was based on previous literature.[14] Clinical department was considered the exposure of interest in these models. Furthermore, effect modification was verified using likelihood ratio tests between models without the interaction term and models with the interaction terms, and a p-value less than 0.05 was considered statistically significant. The interaction terms assessed were those between department and race, department and age, and department and gender. Model calibration was evaluated using a Hosmer-Lemeshow goodness-of-fit test, and model specification was evaluated using a link test. A top box score was defined as a 4 or a 5, whereas a bottom box score was defined as a 1, 2, or 3. This categorization was based on previous literature and past CAHPS guidelines.[15, 16]

Results

15,924 out of 16,850 responses to the telehealth Ambulatory Clinics CAHPS survey were assessed for eligibility in this study. 50 responses from patients visiting small, specialty clinics were excluded, and 1,470 responses from patients visiting clinical departments that made up less than 1% of the patient population were excluded from analysis. 14,404 telehealth patient encounters met the eligibility criteria and were included in this study, as shown in **Figure 1**. **Table 1** describes all participants analyzed throughout this thesis. Telehealth patients were mostly female (64.6%), White (71.5%), Non-Hispanic (83.5%), and English-speaking (98.1%). **Figure 2** maps the number of telehealth visits in the study population between August 2020 and

February 2022 by month. The trend of telehealth and in-person visits during this period was consistent with that of Covid-19 cases and hospitalizations in the greater Sacramento area.[4] The specific results of the logistic regressions for the three telehealth questions examined are as follows:

"*Ease* of talking with the care provider over the video connection." There were 14,066 respondents for this question out of 14,404 total respondents in the study population; the response rate was 97.6%. In addition to the covariates and clinical department, the final logistic regression model created for this outcome also included an interaction term between department and race. A summary of the odds ratios and 95% confidence intervals for the main effects of key variables and covariates for this question are shown in **Table 2.** Notably, the odds of giving a top box score among older patients (25-34, 45-54, 55-64, 65-74, and 75+) were significantly higher than those of the reference category (18-24-year-olds). This model is a good fit for the data [p-value=0.641] and was properly specified according to a link test.

"How well the *video connection* worked during your video visit". 13,670 patients out of 14,404 in the study population responded to this question, for a response rate of 94.9%. The final model for this question included an interaction between department and race, in addition to department and the key covariates. The odds ratios and 95% confidence intervals for the main effects of department and covariates this question can be found in **Table 3**. As with the prior question, individuals aged 25-34, 45-54, 55-64, 65-74, and 75+ had significantly higher odds of giving a top box score than those in the youngest age category. This model is a good fit for the data [p-value=0.772] and was properly specified according to a link test.

"How well the *audio connection* worked during your video visit". 96.9% (13,959 out of 14,404) of participants who visited their provider via telehealth responded to this question. The final logistic regression model for this question included department, key covariates, and the interaction between department and gender. A summary of the odds ratios and 95% confidence intervals for the main effects of clinical department and key covariates for this outcome can be found in **Table 4.** Unlike questions 1 and 2, only the four oldest age categories had significantly higher odds of giving a top box score when compared with the youngest age category. Furthermore, those in the "Asian/Pacific Islander" racial category had 58% increased odds of giving a top box score compared to those in the "White" racial category [OR=1.58; 95% Confidence Interval (1.19, 2.09); p-value=0.001]. This model is a good fit for the data [p-value=0.401] and was properly specified according to a link test.

Conclusion

The objective of <u>Section 1</u> was to determine whether there are differences in patient satisfaction with telehealth visits between clinical departments, age categories, genders, ethnicities, and primary languages using the three telehealth questions on the CAHPS Ambulatory Clinics survey. Throughout all three questions, which related to patient satisfaction with telehealth and the corresponding technology, individuals greater than 24 years of age (ages 25-34, 35-44, 45-54, 55-64, 65-74, and 75+) had higher odds of giving a top box score than those aged 18-24. Though there were some differences in odds of top box scores in other covariates, there does not seem to be any clear differences across the three outcomes in patient satisfaction with telehealth among race, gender, ethnicity, language, and clinical department.

Section 2 – Telehealth vs. In-Person Care

Introduction

As mentioned in Section 1, telehealth has been implemented at many medical centers as a long-term tool for certain visit types and patient groups. Therefore, many researchers are interested in how telehealth compares to in-person care in terms of patient satisfaction.[17, 18] CAHPS is a useful tool in this, as the current version of the survey is used for both telehealth and in-person encounters.[12] Previous studies have evaluated patient satisfaction with different encounter types in the beginning of the pandemic, but they have examined relatively small cohorts.[19, 20] The objective of this section was to compare patient satisfaction scores between in-person and telehealth encounters at a large academic medical center across various departments and specialties serving a diverse patient population.

Methods

This section involves a retrospective, observational analysis of CAHPS patient satisfaction survey scores for patients who met with their care provider either in-person or via telehealth from the Ambulatory Clinics CAHPS survey between August 2020 and February 2022 at 31 specialty clinics and 16 departments from UC Davis Health. Similar to Section 1, all non-surgical, non-pediatric patient encounters were assessed for eligibility and included. Patients visiting their care provider in a surgical clinic, in a pediatric clinic, or those under the age of 18 were not included. Furthermore, small, specialty clinics that made up their own department, or departments that did not comprise 1% or more of the patient population were excluded from analysis. All participants were asked about their in-person or telehealth experience within days of their visit by email, text message, or both. Three attempts were made to survey each patient. The Ambulatory Clinics CAHPS survey had two versions: one that was geared towards telehealth visits and one that was for in-person visits. These surveys differed in question types and wording of similar questions. In the interest of best comparing telehealth and in-person patient experiences, this section analyzes patient survey scores from five questions relating to satisfaction with care providers they met with either in-person or via telehealth, as these questions were identical on both versions of the survey. These questions involved the *explanations* given by the care provider, the *concern* the provider showed, the efforts of the physician to *include* the patients in care decisions, how likely the patient is to *recommend* the care provider, and the *discussion* of proposed treatment. All questions were scored on a 5-point Likert scale, where 1 is "Very Poor" and 5 is "Very Good."

Using Stata 17.0, logistic regression models were constructed including visit type, age, race, gender, ethnicity, language, and department. The inclusion of age, race, gender, ethnicity, and language as potential confounders was confirmed by previous literature.[14] Department's inclusion in the final models was confirmed via likelihood ratio tests and Akaike information criterion (AIC) values. The exposure of interest in these models was visit type—whether a patient encounter occurred in-person or via telehealth. Furthermore, effect modification was verified using likelihood ratio tests between models without interaction term and models with interaction terms, and a p-value less than 0.05 was considered significant. The interaction terms evaluated for each outcome not only included those between covariates and visit type, but also the interaction between department and age, department and gender, and department and race. Model calibration was evaluated using a Hosmer-Lemeshow goodness-of-fit test, and model specification was evaluated using a link test. A top box score defined as a 4 or a 5, whereas a bottom box score was defined as a 1, 2, or 3. As with Section 1, this categorization was based on previous literature and past CAHPS guidelines.[15, 16]

Results

99,514 patient encounters were assessed for eligibility in this section, 15,924 of which were via telehealth and 83,590 were in-person. A total of 8,423 patient encounters—1,520 telehealth encounters and 6,903 in-person encounters—were excluded from analysis. 14,404 (15.8%) telehealth patient encounters and 76,687 (84.2%) in-person patient encounters met the eligibility criteria for this study. The inclusion and exclusion process for this section is detailed in **Figure 1**. Summary demographics for all 91,091 patient visits in the study can be found in **Table 1**. The patient population was mostly White (73.0%), non-Hispanic (85.3%), English-speaking (97.7%), and female (60.7%). This patient demographic breakdown is consistent with previous CAHPS surveys,[21] and is similar to the demographic breakdown of telehealth patients described in Section 1. **Figure 3** depicts all telehealth and in-person encounters by month, showing that the ratio of telehealth to in-person encounters varied over the study period. A summary of odds ratios and 95% confidence intervals for the main effect of telehealth compared to in-person care from the logistic regression output for all five questions can be found in **Table 5**. Specific results for each question are as follows:

"*Explanations* the care provider gave you about your problem or condition." The odds of patients who met with their care provider via telehealth giving a top box score were 2% lower than those of an in-person patient [OR=0.98; 95% Confidence Interval (0.90, 1.08); p-value=0.704] in a model including age, race, gender, ethnicity, language, and department; however, this difference was not statistically significant. However, when including an interaction term between visit type and age in addition to the key covariates, the odds of patients who met with their care provider via telehealth giving a top box score were 57% higher than the odds of patients who met with their provider in-person [OR=1.57; 95% Confidence Interval (0.92, 1.66); p-value=0.096] among those aged 18-24. This relationship was also not significant; there is no significant difference in patient satisfaction between in-person and telehealth encounters for this question. This model is a good fit for the data [p-value=0.671] and was properly specified according to a link test.

"Concern the care provider showed for your questions or worries." The odds ratio for this question was 1.00 [95% Confidence Interval (0.91, 1.11); p-value=0.947] in a model including age, race, gender, ethnicity, language, and department. When an interaction between visit type and race was included, the odds of a telehealth patient giving a top box score were 3% lower than those of an in-person patient giving a top box score [OR=0.97; 95% Confidence Interval (0.86, 1.10); p-value=0.627] in White patients. This is also not a significant association; there is no significant difference in odds of giving a top box score between in-person and telehealth patients for this question. This model is a good fit for the data [p-value=0.961] and was properly specified according to a link test.

"Care provider's efforts to *include* you in decisions about your care." The odds of a telehealth patient giving a top box score were 3% lower than the odds of an in-person patient giving a top box score for this question [OR=0.97; 95% Confidence Interval (0.88, 1.07); p-value=0.598] in a model including age, race, gender, ethnicity, language, and department. However, when including interactions between visit type and race and department and age in addition to key covariates, the odds ratio for this question was 0.94 [95% Confidence Interval

(0.84, 1.05); p-value=0.293] in Whites. This relationship is again not significant. This model is a good fit for the data [p-value=0.780] and was properly specified according to a link test.

"Likelihood of you *recommending* this care provider to others." The odds ratio for this question was 0.97 [95% Confidence Interval (0.89, 1.06); p-value=0.494] in a model including age, race, gender, ethnicity, language, and department. When including interactions between visit type and age, visit type and gender, and department and gender, the odds of a patient visiting their care provider via telehealth giving a top box score were 49% higher than those of a patient visiting their care provider in person [OR=1.49; 95% Confidence Interval (0.86, 2.59); p-value=0.159] among males aged 18-24. This is another insignificant association. This model is a good fit for the data [p-value=0.184] and was properly specified according to a link test.

"Care provider's *discussion* of any proposed treatment (options, risks, benefits, etc.)." The odds ratio for this question was 1.01 [95% Confidence Interval (0.92, 1.10); p-value=0.844] in a model including age, race, gender, ethnicity, language, and department. This is an insignificant association. There was no evidence of effect modification for this question. This model is a good fit for the data [p-value=0.977] and was properly specified according to a link test.

Conclusion

The objective of this section was to assess the comparability of patient satisfaction with their care providers in telehealth and in-person care using CAHPS survey scores. For all five CAHPS questions relating to care providers studied, there was no significant difference in patient satisfaction—measured by the odds of giving a top box score—between in-person and telehealth patient visits when adjusting for covariates. Thus, these methods of care are comparable in terms of patient satisfaction with their care providers.

Section 3 – Top Box Scores

Introduction

AHRQ guidelines on how to analyze CAHPS surveys have changed over time as modifications have been made in the CAHPS survey response values and in questions asked of patients.[13, 16] In the Ambulatory Clinics CAHPS survey, a score of 1 is "Very Poor", 2 is "Poor", 3 is "Fair", 4 is "Good" and 5 is "Very Good" for all questions analyzed in this study. In other versions of CAHPS surveys, such as the 3.1 version which is used for adult medical care, surveys operate on a 4-point scale, where 4 is classified as "Always" and 1 is classified as "Never."[10] Comparisons of results across studies are therefore problematic, as the number of possible options on a Likert scale influences the response of those taking the survey.[22] There are also concerns about the difference between the two top responses, and how this difference is not comparable to other responses, such as the difference between 1 and 2, or between 3 and 4.[23] Some studies consider top box scores to be only the highest score option[24]—in this case, a 5—whereas other studies consider top box scores to be the two highest options on the 5-point scale.[25, 26] The objectives of this section were to determine whether there was a difference in results between the two common categorization of top box scores in literature, and to use a large patient data set to determine which categorization of 5-point Likert-scale scores best captured the data.

Section 3.1 – When a Top Box Score is 5

Methods

The same patient encounter population, inclusion and exclusion criteria, and CAHPS questions analyzed that are used in Section 2 were also used in both Section 3.1 and Section 3.2. For Section 3.1, Stata 17.0 was used to run five logistic regression models—one for each question—that included visit type, age, race, gender, ethnicity, language, and department. The exposure of interest in these models was visit type. Furthermore, effect modification was verified using likelihood ratio tests between models without the interaction terms and models with the interaction terms, and a p-value less than 0.05 was considered significant. The interaction terms evaluated for each outcome not only included those between covariates and visit type, but also the interaction between department and age, department and gender, and department and race. Model calibration was evaluated using a Hosmer-Lemeshow goodness-of-fit test, and model specification was evaluated using a link test. A top box score was considered to be a 5, whereas a bottom box score was considered to be a 1, 2, 3, or 4. This categorization is consistent with current ARHQ guidelines[13] along with previous literature.[25, 26]

Results

A summary of odds ratios and 95% confidence intervals for the main effect of visit type when a top box score is 5 can be found in **Table 6**. Specific results are as follows:

"Explanations the care provider gave you about your problem or condition." The odds ratio for obtaining a top box score for this question was 0.87 [95% Confidence Interval (0.82, 0.91); p-value<0.001] for telehealth visits compared to in-person visits in a model including age, race, gender, ethnicity, language, and department. Similarly, in a model including the

interactions between visit type and language and department and gender, patients visiting their care provider via telehealth have a 14% reduction in odds of giving a top box score relative to patients visiting their care provider in person [OR=0.86; 95% Confidence Interval (0.82, 0.90); p-value<0.001] in English-speaking patients. This is a significant association. This model is a good fit for the data [p-value=0.054] and was properly specified according to a link test.

"Concern the care provider showed for your questions or worries". The odds ratio for giving a top box score in telehealth compared to in-person visits for this question was 0.89 [95% Confidence Interval (0.84, 0.94); p-value<0.001] in a model including age, race, gender, ethnicity, language, and department. After adding the interactions between visit type and race, visit type and department, and department and gender, the odds of a telehealth patient giving a top box score are 9% lower than the odds of an in-person patient giving a top box score [OR=0.91; 95% Confidence Interval (0.80, 1.04); p-value=0.158] among Whites visiting their care provider in internal medicine. This, however, is an insignificant association. This model is a good fit for the data [p-value=0.062] and was properly specified according to a link test.

"Care provider's efforts to *include* you in decisions about your care". The odds ratio for a top box score using telehealth versus in-person care for this question was 0.92 [95% Confidence Interval (0.88, 0.98); p-value<0.001] in a model including age, race, gender, ethnicity, language, and department. After adding the interactions between visit type and ethnicity, visit type and department, and department and gender, there was no significant association between visit type and the odds of giving a top box score among non-Hispanic patients seen in internal medicine [OR=1.00; 95% Confidence Interval (0.88, 1.13); p-value=0.995]. This model is a good fit for the data [p-value=0.585] and was properly specified according to a link test.

“Likelihood of you *recommending* this care provider to others.” The odds ratio for this question was 0.91 [95% Confidence Interval (0.86, 0.96); p-value<0.001] in a model including age, race, gender, ethnicity, language, and department. Similarly, after including the interactions between visit type and race and visit type and language, the odds of a telehealth patient giving a top box score are 14% lower than those of an in-person patient in a model including age, race, gender, ethnicity, language, department, the interaction between visit type and race, and the interaction between visit type and language [OR=0.86; 95% Confidence Interval (0.80, 0.92); p-value<0.001] in White, English-speaking patients. This is a significant association. This model is a good fit for the data [p-value=0.059] and was properly specified according to a link test.

“Care provider’s *discussion* of any proposed treatment (options, risks, benefits, etc.). The odds ratio for a top box score with telehealth versus in-person care for this question was 0.90 [95% Confidence Interval (0.85, 0.94); p-value<0.001] in a model including age, race, gender, ethnicity, language, and department. Results were similar when adding the interaction term between department and gender [OR=0.90; 95% Confidence Interval (0.85, 0.94); p-value<0.001]. This is a significant relationship. This model is a good fit for the data [p-value=0.108] and was properly specified according to a link test.

Conclusion

The objective of this section was to analyze patient satisfaction with their care providers comparing in-person and telehealth patients using a different categorization of top box scores. When top box scores were defined as only a 5 on the 5-point Likert scale, for three of the five CAHPS questions regarding care providers—the *explanations* given by the provider, the likelihood of *recommending* the provider, and the *discussion* of proposed treatments—the odds of a telehealth patient giving a top box score were significantly lower than the odds of an in-person patient giving a top box score in certain subgroups after incorporating effect modifiers. The odds of a telehealth patient giving a top box score were significantly lower than those of an in-person patient for all five outcomes in the general study population in models that included age, race, gender, ethnicity, language, and department. In Section 2, where top box scores were defined as a 4 or a 5, there was no significant difference in odds of giving a top box score between in-person and telehealth patients, but in this section, where top box scores were defined as a 5, there is a significant difference in odds.

Section 3.2 – Model Comparison

Methods

The five care provider CAHPS questions used in Section 2 and Section 3.1 were also used as outcomes for this analysis: the *explanations* given by the care provider, the *concern* they showed, the *inclusion* of the patient in care decisions, the likelihood of the patient to *recommend* the care provider, and the *discussion* of proposed treatment. Using Stata 17.0, binary logistic regressions were run for models without interactions for ease of comparison for what will further be referred to as Outcome A – When a top box score is a “4” or a “5” and Outcome B – When a top box score is a “5”. A link test was still performed to ensure proper model specification. These models were compared against ordinal logistic regression models constructed in Stata for two ordinal outcomes: Outcome C, which treats each score—“1”, “2”, “3”, “4”, and “5”—individually; and Outcome D, which splits the scores into three categories defined by bottom box scores (“1” and “2”), middle box scores (“3” and “4”), and top box scores (“5”). The inclusion of Outcome D was based on AHRQ guidelines.[13] All models included age, gender, race, ethnicity, language, and department, in addition to visit type.

In order to determine which approach to defining an outcome yielded a model that best fit the data, three aspects of each model were examined: model calibration, discrimination, and fitness. Model calibration was defined by a Hosmer-Lemeshow goodness-of-fit test.[27] Measuring the c-statistic, or area under the Receiver Operating Characteristic curve, determined model discrimination due to its ability to measure a logistic regression model's predictive accuracy.[28] Model fit was determined by comparing the pseudo-R^2 values.[29]

Though measuring a c-statistic and Hosmer-Lemeshow goodness-of-fit tests are possible when examining ordinal logistic regression models, these modified values are not comparable to those used in logistic regression models with binary outcomes, so they are not reported for the ordinal models.[30, 31]

Results

"*Explanations* the care provider gave you about your problem or condition."

For Outcome A—the logistic regression model where a top box score is considered a "4" or a "5"—there was no significant association between visit type and odds of giving a top box score after accounting for covariates [OR=0.98; 95% Confidence Interval (0.90, 1.08); p-value=0.704]. This model was properly calibrated according to a Hosmer-Lemeshow goodness-of-fit test [p-value=0.813]. The pseudo-R^2 value for this model was 0.028, and the c-statistic was 0.630.

For Outcome B—the logistic regression model where the top box score is a "5"—the odds of a telehealth patient giving a top box score were 13% lower than the odds of an in-person patient giving a top box score when adjusting for age, race, gender, ethnicity, language, and department [OR=0.87; 95% Confidence Interval (0.82, 0.91); p-value<0.001]. This is a significant association. The Hosmer-Lemeshow goodness-of-fit test confirmed proper model calibration [p-value=0.115]. The c-statistic for this model was 0.568, and the pseudo-R^2 was 0.009. A comparison of the ROC curves for Outcome A and Outcome B can be found in **Figure 4**.

For Outcome C—the ordinal logistic regression model treating each score as an individual category—the odds of a telehealth patient giving a top box score were 13% lower than those of an in-person patient [OR=0.87; 95% Confidence Interval (0.83, 0.92); p-value<0.001] when accounting for covariates. This is a significant relationship. The pseudo-R^2 value for this model was 0.008.

For Outcome D—the ordinal logistic regression model with three outcome categories—the odds of a patient visiting their provider via telehealth giving a top box score were 13% lower than the odds of a patient visiting their provider in-person when accounting for covariates [OR=0.87; 95% Confidence Interval (0.83, 0.92); p-value<0.001]. The pseudo-R^2 for this model was 0.009. A comparison of all models for this CAHPS question can be found in **Table 7.**

"*Concern* the care provider showed for your questions or worries"

For Outcome A there was no significant difference in odds of giving a top box score by visit type [OR=1.00; 95% Confidence Interval (0.91, 1.11); p-value=0.947] when accounting for covariates. This model was properly calibrated according to a Hosmer-Lemeshow goodness-of-fit test [p-value=0.532]. The pseudo-R^2 value for this model was 0.036, and the c-statistic was 0.651.

For Outcome B the odds of a telehealth patient giving a top box score were 11% lower than the odds of an in-person patient giving a top box score when accounting for covariates [OR=0.89; 95% Confidence Interval (0.84, 0.94); p-value<0.001]. This is a significant association. The Hosmer-Lemeshow goodness-of-fit test confirmed proper model calibration [p-value=0.181]. The c-statistic for this model was 0.571, and the pseudo-R^2 was 0.010. A comparison of the ROC curves for Outcome A and Outcome B can be found in **Figure 4**.

For Outcome C the odds of a telehealth patient giving a top box score were 10% lower than those of an in-person patient when accounting for covariates [OR=0.90; 95% Confidence Interval (0.85, 0.95); p-value<0.001]. This is a significant relationship. The pseudo-R^2 value for this model was 0.009.

For Outcome D the odds of a patient visiting their provider via telehealth were 10% lower than the odds of a patient visiting their provider in-person [OR=0.90; 95% Confidence Interval (0.85, 0.95); p-value<0.001] when accounting for covariates. The pseudo-R^2 for this model was 0.010. A comparison of all models for this CAHPS question can be found in **Table 8.**

"Care provider's efforts to *include* you in decisions about your care"

For Outcome A the odds of a patient visiting their care provider via telehealth giving a top box score are 3% lower than the odds of a patient visiting their provider in-person giving a top box score when accounting for covariates [OR=0.97; 95% Confidence Interval (0.88, 1.07); p-value=0.598]. This is an insignificant relationship. This model was properly calibrated according to a Hosmer-Lemeshow goodness-of-fit test [p-value=0.831]. The pseudo-R^2 value for this model was 0.026, and the c-statistic was 0.626.

For Outcome B the odds of a telehealth patient giving a top box score were 8% lower than the odds of an in-person patient giving a top box score when accounting for covariates [OR=0.92; 95% Confidence Interval (0.88, 0.98); p-value<0.001]. This is a significant association. The Hosmer-Lemeshow goodness-of-fit test confirmed proper model calibration [p-value=0.379]. The c-statistic for this model was 0.562, and the pseudo-R^2 was 0.008. A comparison of the ROC curves for Outcome A and Outcome B can be found in **Figure 4**.

For Outcome C the odds of a telehealth patient giving a top box score were 7% lower than those of an in-person patient when accounting for covariates [OR=0.93; 95% Confidence Interval (0.88, 0.98); p-value=0.007]. This is a significant relationship. The pseudo-R^2 value for this model was 0.006.

For Outcome D the odds of a patient visiting their provider via telehealth were 7% lower than the odds of a patient visiting their provider in-person when accounting for covariates [OR=0.93; 95% Confidence Interval (0.88, 0.98); p-value=0.006]. The pseudo-R^2 for this model was 0.007. A comparison of all models for this CAHPS question can be found in **Table 9.**

"Likelihood of you *recommending* this care provider to others"

For Outcome A the odds of a patient visiting their care provider via telehealth giving a top box score are 3% lower than the odds of a patient visiting their provider in-person giving a top box score when accounting for covariates [OR=0.97; 95% Confidence Interval (0.89, 1.06); p-value=0.494]. This is an insignificant relationship.

This model was properly calibrated according to a Hosmer-Lemeshow goodness-of-fit test [p-value=0.169]. The pseudo-R^2 value for this model was 0.028, and the c-statistic was 0.630.

For Outcome B the odds of a telehealth patient giving a top box score were 9% lower than the odds of an in-person patient giving a top box score when accounting for covariates [OR=0.91; 95% Confidence Interval (0.86, 0.96); p-value<0.001]. This is a significant association. The Hosmer-Lemeshow goodness-of-fit test confirmed proper model calibration [p-value=0.228]. The c-statistic for this model was 0.568, and the pseudo-R^2 was 0.010. A comparison of the ROC curves for Outcome A and Outcome B can be found in **Figure 4**.

For Outcome C the odds of a telehealth patient giving a top box score were 9% lower than those of an in-person patient when accounting for covariates [OR=0.91; 95% Confidence Interval (0.86, 0.96); p-value=0.001]. This is a significant relationship. The pseudo-R^2 value for this model was 0.008.

For Outcome D the odds of a patient visiting their provider via telehealth were 9% lower than the odds of a patient visiting their provider in-person when accounting for covariates [OR=0.91; 95% Confidence Interval (0.86, 0.96); p-value=0.001]. The pseudo-R^2 for this model was 0.009. A comparison of all models for this CAHPS question can be found in **Table 10.**

"Care provider's *discussion* of any proposed treatment (options, risks, benefits, etc.)

For Outcome A the odds of a patient visiting their care provider via telehealth giving a top box score are not significantly different than those of an in-person patient when accounting for covariates [OR=1.01; 95% Confidence Interval (0.92, 1.10); p-value=0.844]. This model was properly calibrated according to a Hosmer-Lemeshow goodness-of-fit test [p-value=0.977]. The pseudo-R^2 value for this model was 0.023, and the c-statistic was 0.616.

For Outcome B the odds of a telehealth patient giving a top box score were 10% lower than the odds of an in-person patient giving a top box score when accounting for covariates [OR=0.90; 95% Confidence Interval (0.85, 0.94); p-value<0.001]. This is a significant association. The Hosmer-Lemeshow goodness-of-fit test confirmed proper model calibration [p-value=0.266]. The c-statistic for this model was 0.562, and the pseudo-R^2 was 0.008. A comparison of the ROC curves for Outcome A and Outcome B can be found in **Figure 4**.

For Outcome C the odds of a telehealth patient giving a top box score were 10% lower than those of an in-person patient when accounting for covariates [OR=0.90; 95% Confidence Interval (0.86, 0.95); p-value<0.001]. This is a significant relationship. The pseudo-R^2 value for this model was 0.007.

For Outcome D the odds of a patient visiting their provider via telehealth were 10% lower than the odds of a patient visiting their provider in-person when adjusting for age, race, gender, ethnicity, language, and department [OR=0.90; 95% Confidence Interval (0.85, 0.95); p-value<0.001]. The pseudo-R^2 for this model was 0.008. A comparison of all models for this CAHPS question can be found in **Table 11.**

Conclusion

The objective of this section was to evaluate various approaches of categorizing CAHPS top box scores by examining model fit, calibration, and specification. Outcome A—the logistic regression model where a top box score was a “4” or “5”, which is the most common categorization of top box scores—out-performed the other top box categorization definitions in terms of calibration, discrimination, and fitness. Categorizing both 4’s and 5’s as top box scores best captured the data used in this study which represented a large, diverse patient population at an academic medical center.

Discussion

Patient satisfaction with telehealth examined in this thesis was generally positive and comparable to in-person care. Individuals over the age of 45 had consistently higher odds of giving a top box score than those aged 18-24. An older study indicated that many older individuals have reservations about visiting their care provider via telehealth because they are not comfortable using technology.[32] However, recently older patients' satisfaction with telehealth has been corroborated.[33] Previous studies have found that patients from racial and ethnic minorities were less likely to have visited their care provider via telehealth, and reported difficulty using technology,[34, 35] but there were no notable significant differences in odds ratios across multiple outcomes between races and ethnicities in this study. Past studies have found variable patient satisfaction with their care by department,[36] but this study could not confirm whether these differences exist at UC Davis Health.

Similar to recent findings,[18, 37] patients were as satisfied with telehealth care as they were with in-person care with respect to their care providers. When using both a 4 and 5 on the 5-point Likert scale, there was no significant difference in the odds of a telehealth patient giving a top box score compared to an in-person patient for any of the five provider-oriented CAHPS questions analyzed when adjusting for covariates. However, when using a 5 as the only top box score telehealth patients had significantly lower odds of giving a top box score when compared to an in-person patients when accounting for covariates.

Because the change in categorization of top box scores caused a discrepancy in how the results could be interpreted, examining the categorization that produced the best-fit model became necessary to interpreting the results. When comparing the two aforementioned outcomes to two distinct ordinal outcomes—where one individually categorized Likert scores and the other grouped them into top box, middle box, and bottom box categories—the model that best represented the data was the logistic regression model that considered both a “4” and a “5” on the 5-point Likert scale to be top box scores in the outcome. This was based on model calibration, discrimination, and fitness. Because the 5-point Likert scale is widely used to measure patient satisfaction,[38] it is important to determine which categorization of top box scores yields the best model for the data.

The strengths of this study were that it used a large patient population with a diverse range of medical conditions across multiple clinics and departments at an academic medical center and the utilization of multiple statistical tools to analyze a single data set. The key limitation of this study was that other patient data commonly used in case-mix adjustments, such as socioeconomic status, education level, and general and mental health levels, were not provided and were therefore not included in analysis. An additional limitation was the possibility of the same individual having multiple encounters during the time period of the study, as patient encounters were analyzed rather than individual patients. However, small, specialty clinics were excluded to minimize the chance of counting multiple visits from individual patients.

Future researchers analyzing CAHPS patient satisfaction surveys should carry out a preliminary analysis of the categorization of scores used in their analysis model to ensure they are best capturing their data. Choice of categorization may also be driven by the research question of interest, but researchers should certainly be mindful that different categorizations lead to different results. Furthermore, a qualitative analysis using the comments on CAHPS surveys and similar patient surveys would be helpful in determining if there is a significant difference in patient comments between survey options. Lastly, future studies focusing on departmental differences would provide insight into areas of telehealth and in-person care that may need improvement or other resource allocation.

Appendix

Table 1 – Summary demographics of study sample.

	Telehealth		In-Person		Total	
	N	*(%)*	N	*(%)*	N	*(%)*
	14,404	*(15.8)*	76,687	*(84.2)*	91,091	
Gender						
Male	5,087	*(35.3)*	30,614	*(39.9)*	35,701	*(39.2)*
Female	9,305	*(64.6)*	46,032	*(60.0)*	55,337	*(60.7)*
Other/Declined to State	2	*(<0.01)*	11	*(<0.01)*	13	*(<0.01)*
Missing	10	*(<0.01)*	30	*(<0.01)*	40	*(<0.01)*
Age						
18-24	201	*(1.4)*	853	*(1.1)*	1,054	*(1.2)*
25-34	514	*(3.6)*	2,697	*(3.5)*	3,211	*(3.5)*
35-44	1,097	*(7.6)*	4,361	*(5.7)*	5,458	*(6.0)*
45-54	1,701	*(11.8)*	7,512	*(9.8)*	9,213	*(10.1)*
55-64	3,237	*(22.5)*	15,616	*(20.4)*	18,853	*(20.7)*
65-74	4,944	*(34.3)*	26,994	*(35.2)*	31,938	*(35.1)*
75+	2,709	*(18.8)*	18,640	*(24.3)*	21,349	*(23.4)*
Missing	1	*(<0.01)*	14	*(<0.01)*	15	*(<0.01)*
Race						
White	10,305	*(71.5)*	56,182	*(73.3)*	66,487	*(73.0)*
Black	725	*(5.0)*	3,528	*(4.6)*	4,253	*(4.7)*
Asian/Pacific Islander	937	*(6.5)*	6,011	*(7.8)*	6,948	*(7.6)*
Other/Declined to State	1,859	*(12.9)*	9,285	*(12.1)*	11,144	*(12.2)*
Missing	578	*(4.0)*	1,681	*(2.2)*	2,259	*(2.5)*
Ethnicity						
Non-Hispanic	12,021	*(83.5)*	65,674	*(85.6)*	77,695	*(85.3)*
Hispanic	1,118	*(7.8)*	5,709	*(7.4)*	6,827	*(7.5)*
Other/Declined to State	721	*(5.0)*	3,730	*(4.9)*	4,451	*(4.9)*
Missing	544	*(3.8)*	1,574	*(2.0)*	2,118	*(2.3)*
Language						
English	14,132	*(98.1)*	74,869	*(97.6)*	89,001	*(97.7)*
Other/Declined to State	262	*(1.8)*	1,787	*(2.3)*	2,049	*(2.2)*
Missing	10	*(0.1)*	31	*(0.1)*	41	*(0.1)*
Department						
Internal Medicine	2,717	*(18.9)*	14,913	*(19.4)*	17,630	*(19.4)*
OB/GYN	339	*(2.4)*	2,975	*(3.9)*	3,314	*(3.6)*
Oncology	1,310	*(9.1)*	5,238	*(6.8)*	6,548	*(7.2)*
Nephrology	195	*(1.3)*	611	*(0.8)*	806	*(0.9)*
Allergy/Immunology/Rheum	548	*(3.8)*	1,940	*(2.5)*	2,488	*(2.7)*
Neurology	1,029	*(7.1)*	2,490	*(3.2)*	3,519	*(3.9)*
Hematology	140	*(1.0)*	213	*(0.3)*	353	*(0.4)*
Cardiology	590	*(4.1)*	5,243	*(6.8)*	5,833	*(6.4)*
Family Practice	4,679	*(32.5)*	24,571	*(32.0)*	29,250	*(32.1)*
Endocrinology	581	*(4.0)*	2,019	*(2.6)*	2,600	*(2.8)*
Dermatology	235	*(1.6)*	7,977	*(10.4)*	8,212	*(9.0)*
Gastroenterology	642	*(4.5)*	3,443	*(4.5)*	4,085	*(4.5)*
Pulmonary Medicine	721	*(5.0)*	1,603	*(2.1)*	2,324	*(2.6)*
Anesthesia	188	*(1.3)*	1,503	*(2.0)*	1,691	*(1.9)*
Infectious Diseases	273	*(1.9)*	242	*(0.3)*	515	*(0.6)*
PM&R	217	*(1.5)*	1,706	*(2.2)*	1,923	*(2.1)*

Figure 1 – Consort diagram detailing exclusion criteria and the number of patient encounters excluded, as well as the percentages excluded from their respective populations.

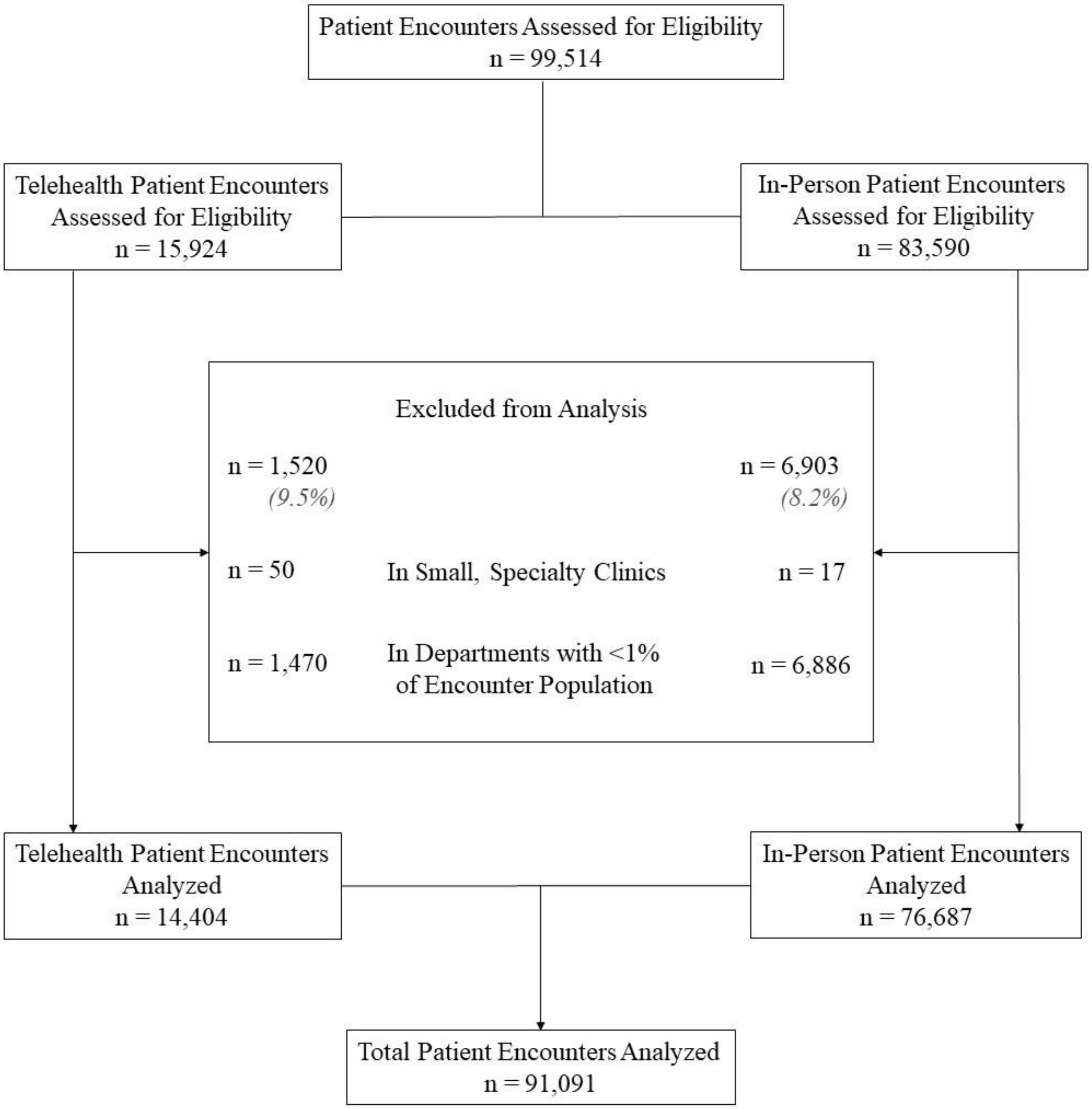


Figure 2 – Total telehealth visits during the study period. All telehealth patient encounters meeting the eligibility criteria for inclusion in the study were counted in this figure.

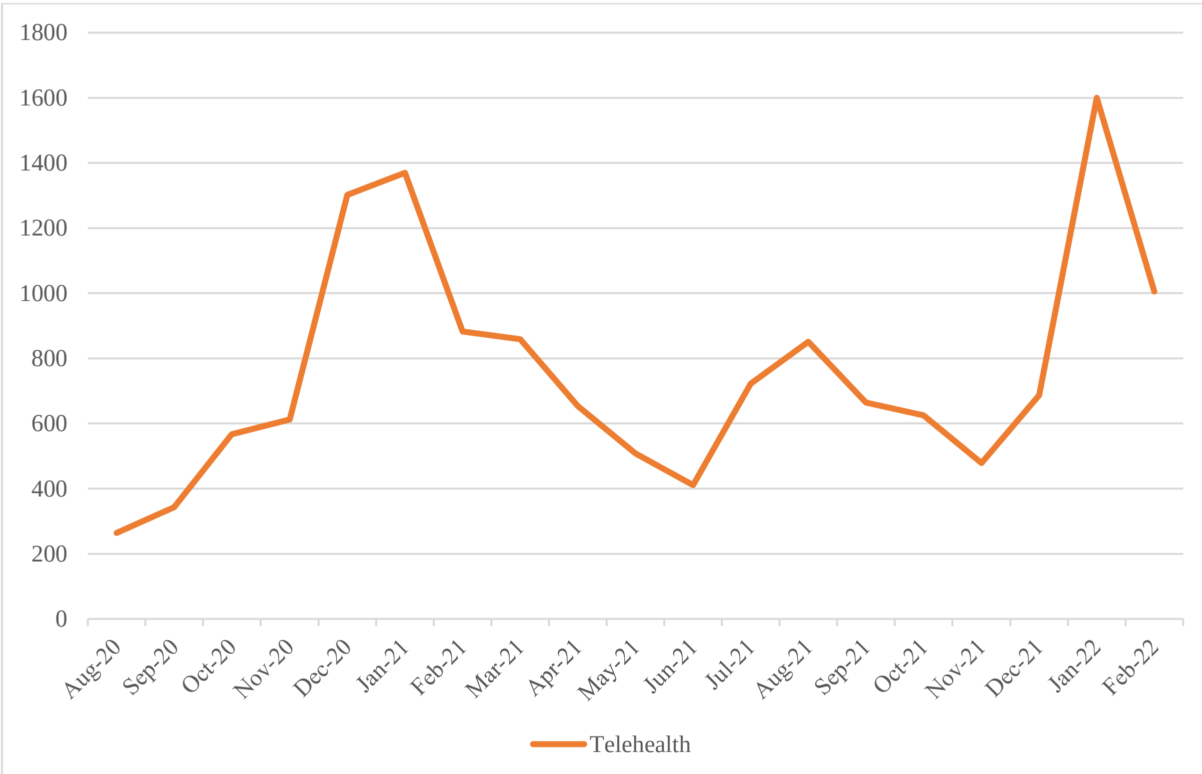


Figure 3 – Total visits by visit type for the study period. All in-person and telehealth patient encounters meeting the eligibility criteria for inclusion in the study were counted in this figure.

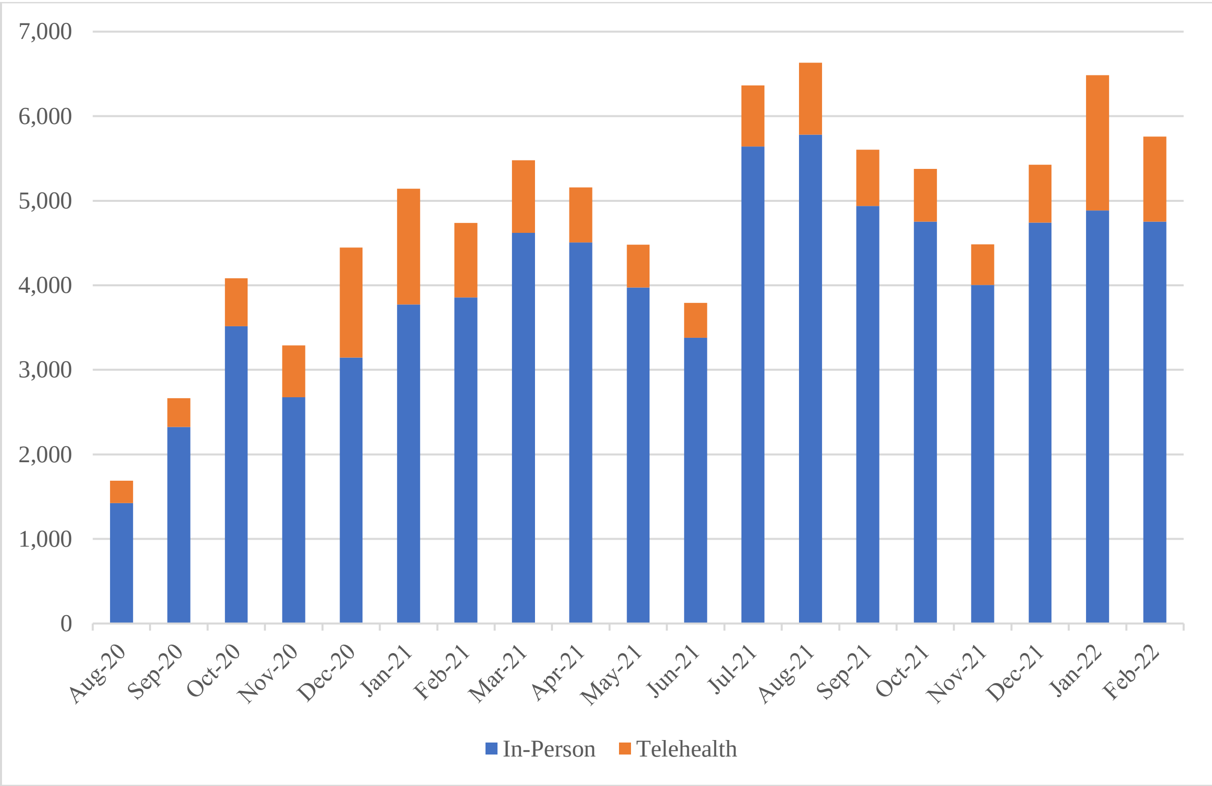


Table 2 – Odds Ratios and 95% Confidence Intervals for Main Effects in the Logistic Regression Model for Telehealth Question 1 – Ease where a Top Box score is a 4 and 5. In addition to main effects, the model also included an interaction between race and department.

	Odds Ratio	95% Confidence Interval	p-value
Gender			
Male	(ref)		
Female	1.03	(0.90, 1.19)	0.662
Other/Declined to State	0.23	(0.01, 4.33)	0.324
Age			
18-24	(ref)		
25-34	1.90	(1.07, 3.40)	0.029
35-44	1.66	(0.99, 2.77)	0.053
45-54	1.82	(1.11, 3.00)	0.018
55-64	1.82	(1.35, 3.49)	0.015
65-74	2.17	(1.35, 3.49)	0.001
75+	2.06	(1.26, 3.36)	0.004
Race*			
White	(ref)		
Black	1.29	(0.56, 2.99)	0.552
Asian/Pacific Islander	1.56	(0.78, 3.12)	0.662
Other/Declined to State	1.81	(0.99, 3.31)	0.055
Ethnicity			
Non-Hispanic	(ref)		
Hispanic	1.01	(0.76, 1.34)	0.923
Other/Declined to State	0.51	(0.38, 0.70)	<0.001
Language			
English	(ref)		
Other/Declined to State	0.87	(0.55, 1.39)	0.570
Department**			
Internal Medicine	(ref)		
OB/GYN	0.72	(0.44, 1.16)	0.175
Oncology	0.94	(0.70, 1.29)	0.722
Nephrology	0.62	(0.36, 1.10)	0.101
Allergy/Immunology/Rheum	1.10	(0.67, 1.82)	0.691
Neurology	0.76	(0.55, 1.06)	0.102
Hematology	2.64	(0.83, 8.42)	0.101
Cardiology	1.00	(0.63, 1.52)	0.993
Family Practice	1.10	(0.87, 1.38)	0.423
Endocrinology	1.00	(0.66, 1.53)	0.986
Dermatology	1.52	(0.70, 3.33)	0.291
Gastroenterology	0.72	(0.49, 1.04)	0.077
Pulmonary Medicine	0.82	(0.56, 1.19)	0.300
Anesthesia	0.74	(0.40, 1.37)	0.340
Infectious Diseases	0.81	(0.47, 1.39)	0.440
PM&R	0.58	(0.33, 0.99)	0.050

*Differences in race are specifically in Internal Medicine Clinics
**Differences between departments are specifically within the White racial group

Table 3 – Odds Ratios and 95% Confidence Intervals for Main Effects in the Logistic Regression Model for Telehealth Question 2 – Video Connection where a Top Box score is a 4 and 5. In addition to main effects, the model also included an interaction between department and race.

	Odds Ratio	95% Confidence Interval	p-value
Gender			
Male	(ref)		
Female	1.03	(0.91, 1.16)	0.609
Other/Declined to State	1	(empty)	
Age			
18-24	(ref)		
25-34	2.21	(1.32, 3.70)	0.003
35-44	1.44	(0.92, 2.24)	0.108
45-54	1.92	(1.24, 2.96)	0.003
55-64	1.97	(1.29, 3.00)	0.002
65-74	1.98	(1.30, 3.00)	0.001
75+	1.96	(1.28, 3.00)	0.002
Race*			
White	(ref)		
Black	1.00	(0.52, 1.89)	0.989
Asian/Pacific Islander	1.57	(0.87, 2.81)	0.132
Other/Declined to State	1.76	(1.04, 2.99)	0.035
Ethnicity			
Non-Hispanic	(ref)		
Hispanic	0.91	(0.72, 1.15)	0.417
Other/Declined to State	0.74	(0.56, 0.99)	0.040
Language			
English	(ref)		
Other/Declined to State	1.06	(0.68, 1.65)	0.798
Department**			
Internal Medicine	(ref)		
OB/GYN	0.78	(0.50, 1.19)	0.245
Oncology	0.76	(0.59, 0.97)	0.029
Nephrology	0.77	(0.46, 1.28)	0.314
Allergy/Immunology/Rheum	0.96	(0.63, 1.43)	0.830
Neurology	0.85	(0.64, 1.14)	0.278
Hematology	2.37	(0.96, 5.88)	0.063
Cardiology	0.79	(0.57, 1.10)	0.169
Family Practice	1.01	(0.84, 1.23)	0.883
Endocrinology	0.99	(0.69, 1.41)	0.954
Dermatology	1.00	(0.57, 1.74)	0.999
Gastroenterology	0.90	(0.64, 1.26)	0.527
Pulmonary Medicine	0.77	(0.57, 1.06)	0.110
Anesthesia	1.10	(0.59, 2.03)	0.765
Infectious Diseases	0.85	(0.53, 1.36)	0.509
PM&R	0.73	(0.43, 1.23)	0.236

*Differences in race are specifically in Internal Medicine Clinics

**Differences between departments are specifically within the White racial group

Table 4 – Odds Ratios and 95% Confidence Intervals for Main Effects in the Logistic Regression Model for Telehealth Question 3 – Audio Connection where a Top Box score is a 4 and 5. In addition to main effects, the model also included an interaction between department and gender.

	Odds Ratio	95% Confidence Interval	p-value
Gender*			
Male	(ref)		
Female	0.95	(0.70, 1.30)	0.627
Other/Declined to State	1	(empty)	
Age			
18-24	(ref)		
25-34	1.51	(0.89, 2.55)	0.123
35-44	1.39	(0.87, 2.23)	0.172
45-54	1.60	(1.01, 2.54)	0.045
55-64	1.66	(1.06, 2.60)	0.026
65-74	1.97	(1.26, 3.08)	0.003
75+	1.92	(1.22, 3.04)	0.005
Race			
White	(ref)		
Black	1.07	(0.81, 1.41)	0.627
Asian/Pacific Islander	1.58	(1.19, 2.09)	0.001
Other/Declined to State	1.03	(0.83, 1.27)	0.807
Ethnicity			
Non-Hispanic	(ref)		
Hispanic	1.00	(0.79, 1.28)	0.973
Other/Declined to State	0.71	(0.54, 0.95)	0.021
Language			
English	(ref)		
Other/Declined to State	0.92	(0.59, 1.42)	0.715
Department**			
Internal Medicine	(ref)		
OB/GYN	0.61	(0.42, 0.88)	0.008
Oncology	0.57	(0.39, 0.83)	0.003
Nephrology	0.57	(0.30, 1.10)	0.094
Allergy/Immunology/Rheum	0.84	(0.43, 1.63)	0.599
Neurology	0.89	(0.56, 1.41)	0.614
Hematology	1.39	(0.42, 4.58)	0.588
Cardiology	0.94	(0.57, 1.57)	0.816
Family Practice	0.96	(0.70, 1.32)	0.806
Endocrinology	1.32	(0.64, 2.70)	0.456
Dermatology	0.98	(0.38, 2.54)	0.970
Gastroenterology	0.53	(0.32, 0.86)	0.011
Pulmonary Medicine	0.61	(0.39, 0.95)	0.030
Anesthesia	0.43	(0.22, 0.84)	0.014
Infectious Diseases	0.83	(0.45, 1.52)	0.541
PM&R	1.07	(0.41, 2.78)	0.888

*Differences in gender are specifically in Internal Medicine Clinics
**Differences between departments are specifically within males

Table 5 – Odds Ratios and 95% Confidence Intervals for the Main Effect of Visit Type in the Logistic Regression Models for all five CAHPS questions relating to care providers examined, where a Top Box score is a 4 and 5 on the Likert Scale. All odds ratios were obtained from logistic regression models that included age, race, gender, ethnicity, language, and department. Models for some questions also included interaction terms (see text for details).

	Odds Ratio	**95% Confidence Interval**	**p-value**
Question 1 – Explanations			
In-Person	(ref)		
Telehealth	1.57	(0.92, 2.66)	0.096
Question 2 – Concern			
In-Person	(ref)		
Telehealth	0.97	(0.86, 1.10)	0.627
Question 3 – Inclusion			
In-Person	(ref)		
Telehealth	0.94	(0.84, 1.05)	0.293
Question 4 – Recommendation			
In-Person	(ref)		
Telehealth	1.49	(0.86, 2.59)	0.159
Question 5 – Discussion			
In-Person	(ref)		
Telehealth	1.01	(0.92, 1.10)	0.844

Table 6 – Odds Ratios and 95% Confidence Intervals for the Main Effect of Visit Type in the Logistic Regression Models for all five CAHPS questions relating to care providers examined, where a Top Box score is a 5 on the Likert Scale. All odds ratios were obtained from logistic regression models that included age, race, gender, ethnicity, language, and department. Models for some questions also included interaction terms (see text for details).

	Odds Ratio	95% Confidence Interval	p-value
Question 1 – Explanations			
In-Person	(ref)		
Telehealth	0.86	(0.82, 0.90)	<0.001
Question 2 – Concern			
In-Person	(ref)		
Telehealth	0.91	(0.80, 1.04)	0.158
Question 3 – Inclusion			
In-Person	(ref)		
Telehealth	1.00	(0.88, 1.13)	0.995
Question 4 – Recommendation			
In-Person	(ref)		
Telehealth	0.86	(0.80, 0.92)	<0.001
Question 5 – Discussion			
In-Person	(ref)		
Telehealth	0.90	(0.85, 0.94)	<0.001

Table 7 – Comparison of Odds Ratios and 95% Confidence Intervals for Visit Type between models including visit type, age, race, ethnicity, gender, language, and department for Question 1 – Explanations, where Outcome A is a logistic regression model with top box (4-5) scores as the outcome of interest, B is a logistic regression model with top box (5) scores as the outcome, C is a ordinal logistic regression model treating each score on the Likert scale as an individual ordinal category, and D is a ordinal logistic regression model with a 3-level ordinal outcome: bottom box (1-2), middle box (3-4), and top box (5).

	Odds Ratio	95% Confidence Interval	p-value	Pseudo R^2
Outcome A				
In-Person	(ref)			
Telehealth	0.98	(0.90, 1.08)	0.704	0.0277
Outcome B				
In-Person	(ref)			
Telehealth	0.87	(0.82, 0.91)	<0.001	0.0093
Outcome C				
In-Person	(ref)			
Telehealth	0.87	(0.83, 0.92)	<0.001	0.0079
Outcome D				
In-Person	(ref)			
Telehealth	0.87	(0.83, 0.92)	<0.001	0.0089

Table 8 – Comparison of Odds Ratios and 95% Confidence Intervals for Visit Type between models including visit type, age, race, ethnicity, gender, language, and department for Question 2 – Concern, where Outcome A is a logistic regression model with top box (4-5) scores as the outcome of interest, B is a logistic regression model with top box (5) scores as the outcome, C is a ordinal logistic regression model treating each score on the Likert scale as an individual ordinal category, and D is a ordinal logistic regression model with a 3-level ordinal outcome: bottom box (1-2), middle box (3-4), and top box (5).

	Odds Ratio	95% Confidence Interval	p-value	Pseudo R^2
Outcome A				
In-Person	(ref)			
Telehealth	1.00	(0.91, 1.11)	0.947	0.0355
Outcome B				
In-Person	(ref)			
Telehealth	0.89	(0.84, 0.94)	<0.001	0.0105
Outcome C				
In-Person	(ref)			
Telehealth	0.90	(0.85, 0.95)	<0.001	0.0090
Outcome D				
In-Person	(ref)			
Telehealth	0.90	(0.85, 0.95)	<0.001	0.0100

Table 9 – Comparison of Odds Ratios and 95% Confidence Intervals for Visit Type between models including visit type, age, race, ethnicity, gender, language, and department for Question 3 – Inclusion, where Outcome A is a logistic regression model with top box (4-5) scores as the outcome of interest, B is a logistic regression model with top box (5) scores as the outcome, C is a ordinal logistic regression model treating each score on the Likert scale as an individual ordinal category, and D is a ordinal logistic regression model with a 3-level ordinal outcome: bottom box (1-2), middle box (3-4), and top box (5).

	Odds Ratio	95% Confidence Interval	p-value	Pseudo R^2
Outcome A				
In-Person	(ref)			
Telehealth	0.97	(0.88, 1.07)	0.598	0.0258
Outcome B				
In-Person	(ref)			
Telehealth	0.92	(0.88, 0.98)	<0.001	0.0075
Outcome C				
In-Person	(ref)			
Telehealth	0.93	(0.88, 0.98)	0.007	0.0064
Outcome D				
In-Person	(ref)			
Telehealth	0.93	(0.88, 0.98)	0.006	0.0071

Table 10 – Comparison of Odds Ratios and 95% Confidence Intervals for Visit Type between models including visit type, age, race, ethnicity, gender, language, and department for Question 4 – Recommendation, where Outcome A is a logistic regression model with top box (4-5) scores as the outcome of interest, B is a logistic regression model with top box (5) scores as the outcome, C is a ordinal logistic regression model treating each score on the Likert scale as an individual ordinal category, and D is a ordinal logistic regression model with a 3-level ordinal outcome: bottom box (1-2), middle box (3-4), and top box (5).

	Odds Ratio	95% Confidence Interval	p-value	Pseudo R^2
Outcome A				
In-Person	(ref)			
Telehealth	0.97	(0.89, 1.06)	0.494	0.0283
Outcome B				
In-Person	(ref)			
Telehealth	0.91	(0.86, 0.96)	<0.001	0.0100
Outcome C				
In-Person	(ref)			
Telehealth	0.91	(0.86, 0.96)	0.001	0.0083
Outcome D				
In-Person	(ref)			
Telehealth	0.91	(0.86, 0.96)	0.001	0.0093

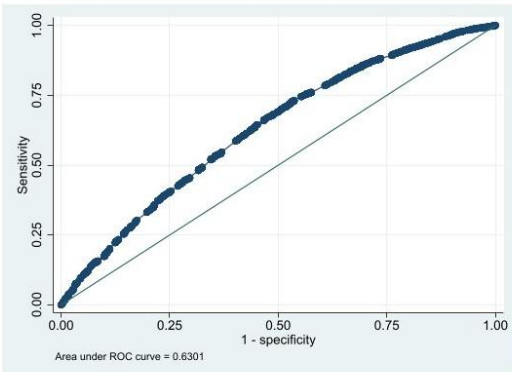
Table 11 – Comparison of Odds Ratios and 95% Confidence Intervals for Visit Type between models including visit type, age, race, ethnicity, gender, language, and department for Question 5 – Discussion, where Outcome A is a logistic regression model with top box (4-5) scores as the outcome of interest, B is a logistic regression model with top box (5) scores as the outcome, C is a ordinal logistic regression model treating each score on the Likert scale as an individual ordinal category, and D is a ordinal logistic regression model with a 3-level ordinal outcome: bottom box (1-2), middle box (3-4), and top box (5).

	Odds Ratio	95% Confidence Interval	p-value	Pseudo R^2
Outcome A				
In-Person	(ref)			
Telehealth	1.01	(0.92, 1.10)	0.844	0.0231
Outcome B				
In-Person	(ref)			
Telehealth	0.90	(0.85, 0.94)	<0.001	0.0079
Outcome C				
In-Person	(ref)			
Telehealth	0.90	(0.86, 0.95)	<0.001	0.0067
Outcome D				
In-Person	(ref)			
Telehealth	0.90	(0.85, 0.95)	<0.001	0.0076

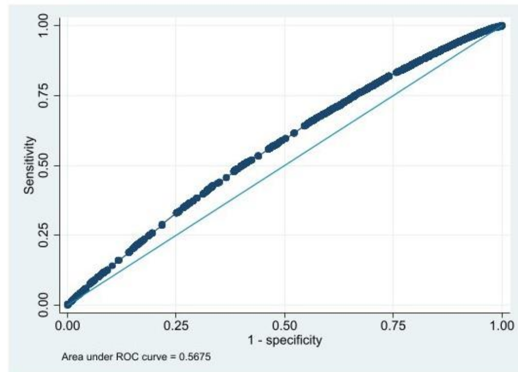
Figure 3 – Area Under the Curve Measurements for all five CAHPS questions comparing Outcome A—when a Top Box score is a "4" or "5"—to Outcome B—when a Top Box score is a "5".

Question 1 - Explanations

Outcome A: Top Box is "4" and "5"

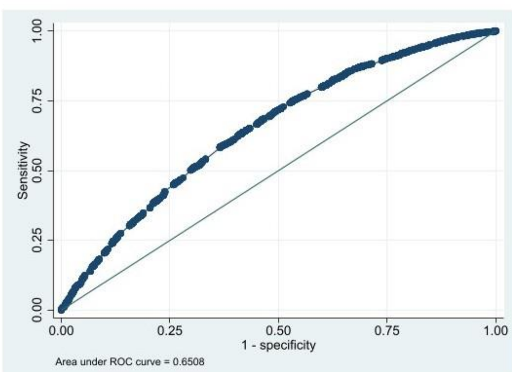


Outcome B: Top Box is "5"

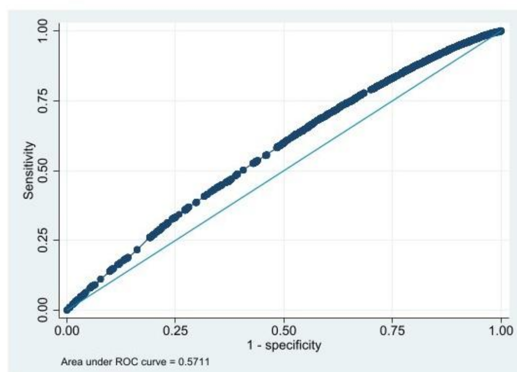


Question 2 - Concern

Outcome A: Top Box is "4" and "5"

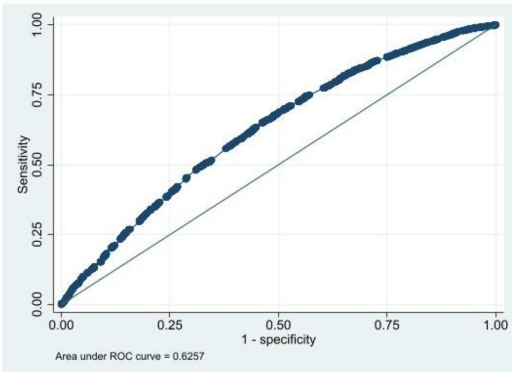


Outcome B: Top Box is "5"

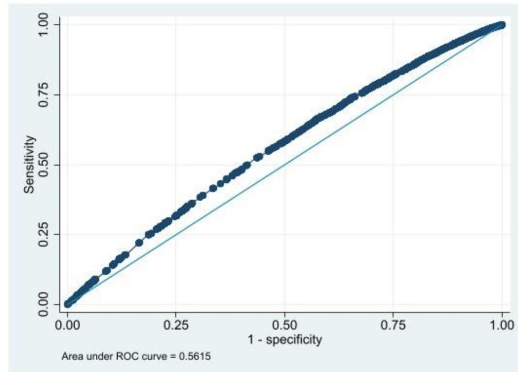


Question 3 - Inclusion

Outcome A: Top Box is “4” and “5”

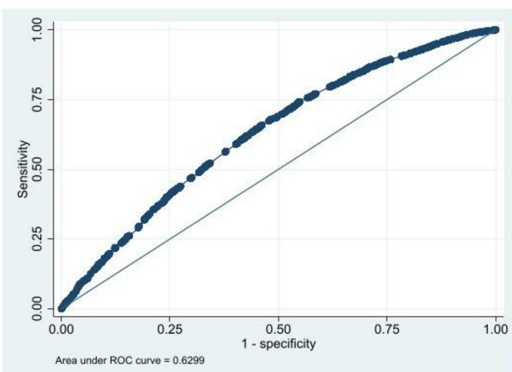


Outcome B: Top Box is “5”

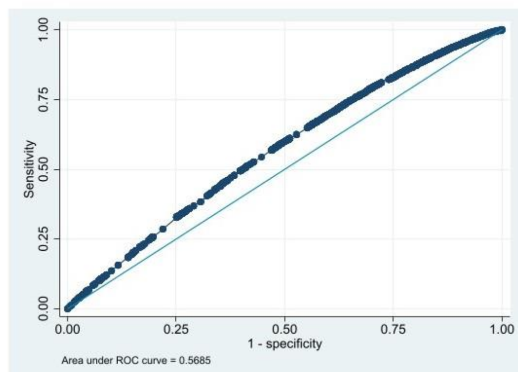


Question 4 - Recommendation

Outcome A: Top Box is “4” and “5”

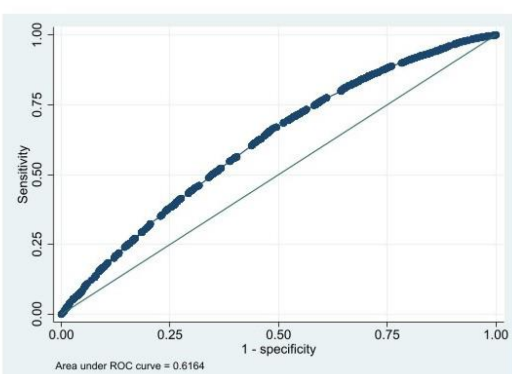


Outcome B: Top Box is “5”

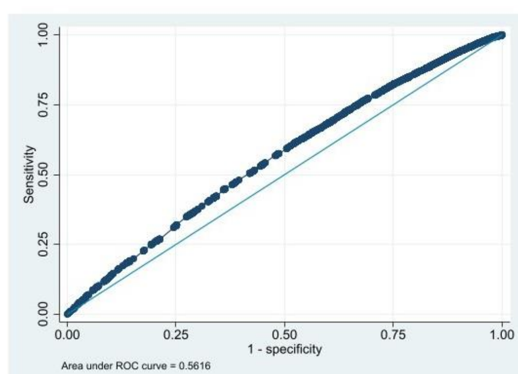


Question 5 - Discussion

Outcome A: Top Box is “4” and “5”



Outcome B: Top Box is “5”



References

1. Catalyst NEJoM. What Is Telehealth? New England Journal of Medicine. 2019.

2. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly T, et al. Trends in the Use of Telehealth During the Emergence of the COVID-19 Pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep. 2020;69(43):1595-9.

3. Wong MYZ, Gunasekeran DV, Nusinovici S, Sabanayagam C, Yeo KK, Cheng CY, et al. Telehealth Demand Trends During the COVID-19 Pandemic in the Top 50 Most Affected Countries: Infodemiological Evaluation. JMIR Public Health Surveill. 2021;7(2):e24445.

4. WHO Coronavirus (COVID-19) Dashboard [Internet]. World Health Organization. 2022. Available from: https://covid19.who.int/more-resources.

5. Ackerman SL, Gleason N, Shipman SA. Comparing Patients' Experiences with Electronic and Traditional Consultation: Results from a Multisite Survey. J Gen Intern Med. 2020;35(4):1135-42.

6. Hays RD, Shaul JA, Valerie SLW, Lubalin JS, Harris-Kojetin LD, Sweeny SF, et al. Psychometric Properties of the CAHPS™ 1.0 Survey Measures. Medical Care. 1999;37(3):MS22-MS31.

7. Darby C Fau - Hays RD, Hays Rd Fau - Kletke P, Kletke P. Development and evaluation of the CAHPS hospital survey. (0017-9124 (Print)).

8. Quigley DD, Mendel PJ, Predmore ZS, Chen AY, Hays RD. Use of CAHPS(®) patient experience survey data as part of a patient-centered medical home quality improvement initiative. (1179-3201 (Print)).

9. Schmocker RK, Cherney Stafford LM, Siy AB, Leverson GE, Winslow ER. Understanding the determinants of patient satisfaction with surgical care using the Consumer Assessment of Healthcare Providers and Systems surgical care survey (S-CAHPS). Surgery. 2015;158(6):1724-33.

10. Quality AfHRa. CAHPS Clinician & Group Survey Version 3.1 [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2022 [Available from: https://www.ahrq.gov/sites/default/files/wysiwyg/cahps/surveys-guidance/cg/adult-english-cg-3-1-2351a.pdf

11. Chao GF, Li KY, Zhu Z, McCullough J, Thompson M, Claflin J, et al. Use of Telehealth by Surgical Specialties During the COVID-19 Pandemic. JAMA Surgery. 2021;156(7):620-6.

12. Quality AfHRa. CAHPS Clinician & Group Survey [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2022 [Available from: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html

13. Quality AfHRa. AHRQ Publishing and Communications Guidelines [Internet]. Rockville, MD: Agency for Healthcare Research and Quality; 2022 [Available from: https://www.ahrq.gov/research/publications/pubcomguide/index.html

14. O'Malley AJ, Zaslavsky AM, Elliott MN, Zaborski L, Cleary PD. Case-mix adjustment of the CAHPS Hospital Survey. Health Serv Res. 2005;40(6 Pt 2):2162-81.

15. Dad T, Tighiouart H, Lacson E, Jr., Meyer KB, Weiner DE, Richardson MM. Hemodialysis patient characteristics associated with better experience as measured by the In-center Hemodialysis Consumer Assessment of Healthcare Providers and Systems (ICH CAHPS) survey. BMC Nephrol. 2018;19(1):340.

16. 2012 Chartbook: What Patients Say About Their Health Care Providers and Clinics [Internet]. Agency for Healthcare Research and Quality. 2013.

17. Bilimoria KY, Zhan T, Durst DA, Merkow RP, Sama PR, Bahaveolos SA, et al. Comparison of Patient Experience with Telehealth vs. In-Person Visits Before and During the COVID-19 Pandemic. Jt Comm J Qual Patient Saf. 2021;47(8):533-6.

18. Hays RD, Skootsky SA. Patient Experience with In-Person and Telehealth Visits Before and During the COVID-19 Pandemic at a Large Integrated Health System in the United States. J Gen Intern Med. 2022;37(4):847-52.

19. Holzman SA, Davis-Dao CA, Khoury AE, Fortier MA, Z NK. Telemedicine and patient satisfaction during the COVID-19 pandemic: A case-control study of outpatient pediatric urology patients. J Child Health Care. 2021:13674935211058272.

20. Levine S, Gupta R, Alkwatli K, Almoushref A, Cherian S, Jimenez DF, et al. Telehealth Perceptions Among US Immigrant Patients at an Academic Internal Medicine Practice: Cross-sectional Study. JMIR Hum Factors. 2022;9(3):e36069.

21. Demographic Characteristics of SEER-CAHPS Respondents by Health Plan Type [Internet]. National Institute of Health. 2022. Available from: https://healthcaredelivery.cancer.gov/seer-cahps/aboutdata/demographics.html. .

22. Simms LJ, Zelazny K, Williams TF, Bernstein L. Does the number of response options matter? Psychometric perspectives using personality questionnaire data. Psychol Assess. 2019;31(4):557-66.

23. Weijters B, Millet K, Cabooter E. Extremity in horizontal and vertical Likert scale format responses. Some evidence on how visual distance between response categories influences extreme responding. International Journal of Research in Marketing. 2021;38(1):85-103.

24. Schmocker RK, Cherney Stafford LM, Winslow ER. Satisfaction with surgeon care as measured by the Surgery-CAHPS survey is not related to NSQIP outcomes. Surgery. 2019;165(3):510-5.

25. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, et al. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. Health Serv Res. 2009;44(2 Pt 1):501-18.

26. Fareed N, MacEwan SR, Vink S, Jonnalagadda P, McAlearney AS. Relationships between patient portal activation and patient satisfaction scores among CG-CAHPS and HCAHPS respondents. Am J Manag Care. 2022;28(1):25-31.

27. Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A Comparison of Goodness-of-Fit Tests for the Logistic Regression Model. Statistics in Medicine. 1997;16(9):965-80.

28. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. BMC Med Res Methodol. 2012;12:82.

29. Hemmert GAJ, Schons LM, Wieseke J, Schimmelpfennig H. Log-likelihood-based Pseudo-R2 in Logistic Regression:Deriving Sample-sensitive Benchmarks. Sociological Methods & Research. 2018;47(3):507-31.

30. Fagerland M, Hosmer D. How to Test for Goodness of Fit in Ordinal Logistic Regression Models. The Stata Journal: Promoting communications on statistics and Stata. 2017;17:668-86.

31. Waegeman W, Baets BD, Boullart L. ROC analysis in ordinal regression learning. Pattern Recogn Lett. 2008;29(1):1–9.

32. Negrini S, Kiekens C, Bernetti A, Capecci M, Ceravolo MG, Lavezzi S, et al. Telemedicine from research to practice during the pandemic. "Instant paper from the field" on rehabilitation answers to the COVID-19 emergency. Eur J Phys Rehabil Med. 2020;56(3):327-30.

33. Alsabeeha NHM, Atieh MA, Balakrishnan MS. Older Adults' Satisfaction with Telemedicine During the COVID-19 Pandemic: A Systematic Review. Telemed J E Health. 2022.

34. Albon D, Van Citters AD, Ong T, Dieni O, Dowd C, Willis A, et al. Telehealth use in cystic fibrosis during COVID-19: Association with race, ethnicity, and socioeconomic factors. Journal of Cystic Fibrosis. 2021;20:49-54.

35. Jewett PI, Vogel RI, Ghebre R, Hui JYC, Parsons HM, Rao A, et al. Telehealth in cancer care during COVID-19: disparities by age, race/ethnicity, and residential status. Journal of Cancer Survivorship. 2022;16(1):44-51.

36. Maher DP, Hess D, Edwards C, Allen L. Changes in Patient Satisfaction Scores During the Early COVID-19 Pandemic. J Patient Exp. 2021;8:23743735211034610.

37. Rodrigues A, Yu JS, Bhambhvani H, Uppstrom T, Ricci WM, Dines JS, et al. Patient Experience and Satisfaction with Telemedicine During Coronavirus Disease 2019: A Multi-Institution Experience. Telemed J E Health. 2022;28(2):150-7.

38. Taylor C, Benger JR. Patient satisfaction in emergency medicine. Emerg Med J. 2004;21(5):528-32.