Using Patient Experience Surveys to Identify Potential Diagnostic Safety Breakdowns: A Mixed Methods Study

Kelley M. Baker, MA[*], **Mark Brahier, MD**[†,‡], **Mara Penne, MD**[†,§], **Mary A. Hill, MS**[‖,¶], **Siara Davis, BS**[**,††], **William J. Gallagher, MD**[‡‡], **Kristen E. Miller, DrPH**[§§], **Kelly M. Smith, PhD**[‖,¶]

[*]MedStar Health Research Institute, Washington, District of Columbia

[†]Georgetown University School of Medicine, Washington, District of Columbia (former)

[‡]Division of General Internal Medicine, Duke University School of Medicine, Durham, North Carolina (current)

[§]Department of Ophthalmology, Wills Eye Hospital, Philadelphia, Pennsylvania (current)

[‖]Institute for Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[¶]Department of Patient-Oriented Research, Michael Garron Hospital, Toronto, Ontario, Canada

[**]Howard University, Washington, District of Columbia (former)

[††]Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (current)

[‡‡]Department of Family Medicine, Georgetown University School of Medicine, Washington, District of Columbia

[§§]National Center for Human Factors in Healthcare, MedStar Health Research Institute, Washington, District of Columbia.

Abstract

Objectives: One in 20 outpatients in the United States experiences a diagnostic error each year, but there are no validated methods for collecting feedback from patients on diagnostic safety. We examined patient experience surveys to determine whether patients' free text comments indicated diagnostic breakdowns. Our objective was to evaluate associations between patient-perceived diagnostic breakdowns reported in free text comments and patients' responses to structured survey questions.

Methods: We conducted an exploratory mixed methods study using data from patient experience surveys collected from adult ambulatory care patients March 2020 to June 2020 in a large U.S.

Correspondence: Kelley M. Baker, MA, MedStar Health Research Institute, 3007 Tilden Street NW, Suite 6N, Washington, DC 20008, (kelley.m.baker@medstar.net).

The authors disclose no conflict of interest.

health system. Data analysis included content analysis of qualitative data and statistical analysis of quantitative data.

Results: In 2525 surveys with negative comments, 619 patients (24.5%) identified diagnostic breakdowns, including issues with accuracy (n = 282, 46%), timeliness (n = 243, 39%), or communication (n = 290, 47%); some patients (n = 181) reported breakdowns in multiple categories. Patients who gave a low average score (50 or less on a 100-point scale) on provider questions were almost seven times more likely to perceive a diagnostic breakdown than patients who scored their provider higher. Similarly, patients who gave a low average score on practice-related questions were twice as likely to perceive a diagnostic breakdown.

Conclusions: Patient feedback in routinely collected patient experience surveys is a valuable and actionable information source on diagnostic breakdowns in the ambulatory setting. The more easily monitored structured survey data provide a screening method to identify encounters that may have included a patient-perceived diagnostic breakdown and therefore require further examination.

Keywords

Diagnostic error is the "failure to establish an accurate and timely explanation of the patient's health problem or to communicate that health problem to the patient,"[1] and most people will experience a diagnostic error in their lifetime.[1] The problem is especially acute in ambulatory care, where 1 in 20 outpatients experiences a diagnostic error each year.[2] However, few health care organizations routinely measure diagnostic error.[3] Historically, diagnostic errors were recognized when an autopsy or malpractice claim highlighted the error. More recently, diagnostic error measurement has become more proactive,[4] but it remains challenging.[3]

Patients and families are a reliable source of relevant safety concerns,[5–12] including breakdowns not reported or detected elsewhere,[13–16] and documentation errors that could affect diagnostic reasoning.[17] There are, however, no validated methods for collecting patient feedback on diagnostic error.[18] Patient complaints are the most common sources of patient and family input. There is evidence of the benefit of mining patient complaints for medical error in general,[19–22] and diagnostic error in particular,[7,23] but there are also drawbacks. Patient complaints are voluminous and diverse, ranging from experience complaints like insufficient parking or food quality to more serious complaints of medical error or harm.[22] It is difficult to find the signal of diagnostic error in the volume of patient complaints.

Patient-reported experience measures (PREMs) provide routinely collected and readily available patient feedback. In the ambulatory care setting in the United States, PREMs often come from the CAHPS (Consumer Assessment of Healthcare Providers and Systems) survey. Typically, CAHPS surveys include structured data questions in the form of yes/no questions, such as "Did someone from this provider's office follow up to give you your test results?" and Likert scale questions, such as "Using any number from 0 to 10, where

0 is the worst visit possible and 10 is the best visit possible, what number would you use to rate your most recent visit?"[24] CAHPS surveys may also include free text questions to allow respondents to provide narrative feedback. The free text answers offer information on the patient experience that structured data alone cannot.[25] However, much like patient complaints, patient experience comments are abundant and diverse, and reviewing them to identify diagnostic breakdowns would be time consuming, labor intensive, and require subject matter expertise on diagnosis. There is little evidence that patient experience surveys are being used to detect diagnostic safety issues.

In this study, we evaluated CAHPS surveys with free text comments to determine whether the comments indicated patient-perceived diagnostic breakdowns in accuracy (missed diagnosis or wrong diagnosis), timeliness (delay in diagnosis or care), or communication during ambulatory care encounters. We then used the structured data fields from the surveys to predict which patients were more likely to report a perceived diagnostic breakdown. The objective of the study was to evaluate the association between patient-perceived diagnostic breakdowns reported in free text comments and patient responses to structured survey questions.

METHODS

Study Design and Setting

We used an exploratory mixed methods study design to investigate patient-perceived diagnostic breakdowns during ambulatory care visits. Data for the study were collected March 2020 to June 2020 in the ambulatory care arm of a large, distributed health care system in the Mid-Atlantic United States. Ambulatory encounters included visits with primary care providers and specialists; the data did not include urgent care and emergency department encounters. The research was approved by the local institutional review board. This paper is written following the Strengthening the Reporting of Observational Studies in Epidemiology guidelines.[26]

Data Source

Data for the study were obtained from a proprietary vendor-delivered CAHPS survey (Press Ganey Medical Practice Survey), a commonly used PREM of outpatient experience. Patients were randomly selected to be surveyed by the vendor within 2 weeks of an ambulatory appointment. The survey originally included six domain areas: access, moving through your visit, nurse/assistant, care provider, personal issues, and overall (a practice-related assessment). A seventh domain area for telemedicine visits was added during the data collection period, on May 30, 2020. Within each topic area, there were 2–6 structured questions with 5-point Likert scale responses. In addition, for each topic, there was an unstructured free text question asking the patient to describe their good or bad experience. The vendor sentiment coded the free text comments as positive, negative, mixed, or unknown before providing the survey data to the health system.

Participants

Participants were adult patients who returned a survey between March 2020 and June 2020 following a confirmed ambulatory encounter. Patients were excluded from the study if they did not provide free text comments or if their free text comments were sentiment coded as positive.

Coding Process

We systematically coded free text responses using content analysis.[27] First, the sentiment coding was confirmed. Comments that were sentiment coded as negative but found to be positive or neutral were excluded. For example, the comment "I couldn't ask for a better doctor" might be initially sentiment coded as negative by the vendor but revised to be coded positive by the research team. Next, all confirmed negative comments were reviewed to identify patient-perceived diagnostic breakdowns. Based on the definition of diagnostic error from the National Academies of Sciences, Engineering, and Medicine,[1] we coded breakdowns in diagnostic accuracy, timeliness, or communication.

The coding was led by 2 senior researchers and completed by a team of 4 research assistants. To achieve consistent coding, 30 cases were selected for training. An additional 10 cases were then randomly selected and coded by all coders to confirm multicoder reliability using the Brennan and Prediger coefficient, which is appropriate for data with a skewed distribution,[28] as was the case for these data. Multicoder reliability statistics were calculated using Stata 16 (StataCorp, College Station, TX), and an intercoder reliability of 0.76 (95% confidence interval [CI] 0.67, 0.86) was achieved, which is considered "good" agreement.[29] During case coding, coders were encouraged to identify cases that warranted further discussion. Such cases were debriefed weekly, and coding validated by the senior researchers. Study data were imported to and managed using REDCap electronic data capture tools (REDCap, Vanderbilt University Medical Center, Nashville, TN).[30,31] REDCap is a secure, web-based software platform designed to support data capture for research studies.

Variables

The primary outcome measures for the study were the coded diagnostic breakdowns in accuracy, timeliness, or communication (yes/no). Accuracy breakdowns were defined as a wrong or missed diagnosis or cases when a patient questioned the providers' ability to accurately diagnose their condition. Timeliness breakdowns included cases where diagnosis or care was delayed or did not occur. Communication breakdowns included communication issues either among the care team or between the patient and provider. Examples of patient-provider breakdowns included the patient not feeling heard by the provider, the patient desiring more or better communication, and language barriers.

Breakdowns were only coded for the encounter associated with the survey. For example, a patient might comment, "Sometimes it's hard to get an appointment." While this suggests an issue with diagnostic timeliness in general, it is not specific to the index encounter, so it was not coded as a breakdown. In addition, inconvenience was not coded as a breakdown unless the patient specifically indicated that it impacted diagnosis. "I had to wait 2 hours"

was considered an inconvenience, while "I appreciated the referral to a podiatrist but concerned it took so long and my foot health suffered" was a breakdown in timeliness. Similarly, telehealth technical difficulties were not coded as diagnostic breakdowns unless they specifically affected diagnosis. "The video didn't work, so the doctor couldn't see my rash" was coded as a diagnostic breakdown in accuracy, but "We couldn't get the video to work, so we had a phone call" was not.

The structured response data in the dataset consisted of patient responses to 5-point Likert scale questions, where scores ranged from very poor (0) to very good (100). For analysis purposes, average scores were calculated across all questions for each topic area and designated as low average score (0 to 50) or high average score (greater than 50). Other variables in the dataset included characteristics of the patient (age, sex, race), appointment (telehealth or in-person), and survey (paper or Internet survey, elapsed time between appointment and survey return).

With the exception of several structured survey questions, missing data were less than 11% for all variables. The moving through your visit questions (2), nurse/assistant questions (2), telemedicine questions (3), and two of the personal issues questions had missing data greater than 40%. These 9 questions were excluded from further analysis.

Data Analysis

Data were summarized using simple descriptive statistics, including frequencies and counts for categorical variables and means and standard deviations for continuous variables. Bivariate analyses were conducted using t tests for continuous variables and χ^2 for categorical data. Statistical tests were 2-sided using $\alpha = 0.05$, and the Benjamini-Hochberg correction was applied to control the false discovery rate for multiple comparisons.[32]

Multivariable logistic regression was used to determine if low average scores on access, provider, personal issues, and practice-related questions were predictive of patient-perceived diagnostic breakdowns. Patient, appointment, and survey characteristics were included in the regression models when bivariate analysis indicated some association between the characteristic and the outcome ($P < 0.25$). Regression models were run for any diagnostic breakdown and each of accuracy, timeliness, and communication breakdowns. Results are reported as both unadjusted and adjusted odds ratios with 95% CIs, and an alpha of 0.05 was considered significant. All analyses were conducted using Stata 16 (StataCorp, College Station, TX).[29]

RESULTS

Participants

During the study period, there were a total of 11,449 unique patient encounters where the patient provided free text comments. Of the 11,449 comments, 3607 were sentiment coded as negative by the vendor. Of those, 1082 comments were excluded from further analysis because the patient did not have an ambulatory encounter (n = 17), the patient was younger than 18 (n = 109), or the comments were recoded as positive or neutral (n = 956), leaving a final dataset of 2525 confirmed negative comments (Fig. 1). The participants who provided

confirmed negative comments had an average age of 62.3 years (SD = 14.3) and were majority female (n = 1644; 65.1%) and White (n = 1588; 68.5%) (Table 1). Most had a telehealth visit (n = 1427; 60.5%) and completed their survey via the Internet (n = 1928; 76.4%). The average elapsed time from encounter to survey response was 15.8 days (SD = 17.7). One in 4 patients reported at least one diagnostic breakdown (n = 619; 24.5%).

Patient-Perceived Diagnostic Breakdowns

Within the 619 encounters with patient-perceived breakdowns, there were a total of 815 breakdowns, categorized as issues with accuracy (n = 282, 46%), timeliness (n = 243, 39%), or communication (n = 290, 47%). Some participants (n = 181, 29%) identified breakdowns in more than 1 category. Examples of accuracy breakdowns included:

> "The doctor looked at my finger for one second and said I would be alright. It is still hurting." – 57-year-old female (Participant ID 2755)

> "Due to the telemed appointment, the doctor could not physically diagnose my symptoms." – 69-year-old male (Participant ID 528)

For breakdowns in timeliness, patients stated, for example:

> "Because the MRI broke after they put my father through it twice the first time, they were unable to squeeze us in for a week and a half to 2 weeks to redo it." – 87-year-old male (Participant ID 3463)

> "Still waiting for a call to schedule surgery." – 70-year-old female (Participant ID 3240)

Examples of communication breakdowns included:

> "Dr. X was not listening to me…I kept telling him how much pain I was in from my shoulder, and he just dismissed me." – 40-year-old female (Participant ID 1507)

> "There is no way that my mom will have access to quality care without advocacy from me or an interpreter for her… Broken English should not be the basis for if an individual understands." – 66-year-old female (Participant ID 2950)

> "I would just suggest the doctor take a few minutes to explain things to the patient better." – 38-year-old female (Participant ID 621)

> "There needs to be more communication with other doctor which is very important to me." – 69-year-old female (Participant ID 560)

Association Between Perceived Diagnostic Breakdowns and Patient, Appointment, and Survey Characteristics

Bivariate analyses found that a significantly greater percentage of patients who responded by paper survey reported a diagnostic breakdown compared with those patients who responded by Internet survey (Table 1). In addition, a longer elapsed time between the appointment date and the survey response was associated with perceiving a diagnostic breakdown.

Structured Survey Responses

Twelve structured response questions, consisting of questions on access (3 questions), the provider (6 questions), personal issues (1 question), and the overall practice (2 questions), were analyzed. Low average scores (0–50) were reported by 10%–35% of patients across the questions (Table 2). For all questions, χ^2 analyses revealed that there was a statistically significant association between providing a low score and perceiving a diagnostic breakdown.

Association Between Perceived Diagnostic Breakdowns and Structured Survey Responses

Multivariable logistic regression results indicated that, after controlling for patient (age, sex), appointment (telehealth or in-person), and survey (paper or Internet survey, elapsed time between appointment and survey return) characteristics, low average scores on the questions in the provider and practice-related domains were predictive of patient-perceived diagnostic breakdowns (Table 3). Specifically, a patient who gave a low average score on the provider questions was almost seven times more likely to perceive a diagnostic breakdown (aOR 6.8, 95% CI 4.7, 10.0) compared with a patient who gave a high average provider score. A patient who scored the practice-related questions with a low average score was twice as likely to perceive a diagnostic breakdown (aOR 2.3, 95% CI 1.7, 3.1) as a patient who provided a high average practice-related score.

Results varied slightly by breakdown type (Table 4). Patients who provided low average scores on provider questions were more likely to perceive breakdowns in accuracy and communication, but not timeliness. Patients who gave low average scores on access questions were more likely to perceive timeliness breakdowns, but less likely to perceive accuracy issues. Low average scores on practice-related questions were significantly associated with all 3 breakdown types.

DISCUSSION

This study demonstrates that patient feedback in routinely collected PREMs is a valuable and actionable information source on diagnostic breakdowns in the ambulatory setting. First, we established that patients do perceive diagnostic breakdowns and do report the breakdowns through their free text comments on patient experience surveys. Among those patients who provided negative free text comments, nearly one quarter revealed a perceived breakdown in the diagnostic process. The breakdowns were relatively evenly distributed across accuracy, timeliness, and communication issues, which is consistent with prior research on the nature of diagnostic breakdowns.[33] Next, we showed that low scores on some structured survey questions were significantly associated with perceived diagnostic breakdowns. Those patients who provided an average score of less than 50 on a 100-point scale on the provider and practice-related questions were significantly more likely to write comments indicating a perceived diagnostic breakdown.

This study highlights the importance of the provider's role in how the patient perceives the quality of their diagnostic process. The provider questions on the survey include the

concepts of concern the provider showed for the patient's questions or worries, extent to which the provider listened, and explanations the provider gave. Similar themes are found consistently in the diagnostic error literature. Patients believe that diagnostic breakdowns occur when providers do not listen to them, are dismissive of their concerns, or are not aligned on the significance of their symptoms.[34–36] It follows that low scores on the provider questions could indicate perceived issues in the diagnostic process.

Interestingly, a longer average elapsed time between appointment date and survey response was associated with increased likelihood of reporting a diagnostic breakdown. This may be because some diagnostic breakdowns do not become apparent until time has passed. For example, if the patient receives a different diagnosis weeks or months after their initial encounter, they may come to believe the second diagnosis is the correct diagnosis and recognize the initial diagnosis as an error. There is evidence that dissatisfied patients do not tend to voice their complaint at the time of service but later wish they had.[37] This may be true with PREM surveys too. Those patients who respond soon after the encounter may not express dissatisfaction, while those who wait may find their dissatisfaction grow to the point that they are willing to report it.

There are currently no PREMs focused specifically on diagnostic accuracy, timeliness, and communication, although the CAHPS survey can be customized to include supplemental questions that may provide insight into the diagnostic process.[24] In the absence of a diagnostic safety-specific PREMs survey or diagnostic safety-specific questions within a PREMs survey, this study provides proof of concept for a quality improvement method to fill the diagnostic safety measurement gap. It uses PREMs data that are already being collected and does not require additional data collection. Because it is easier to monitor structured numeric data, ambulatory practices could flag surveys with low scores, particularly on the provider and practice-related questions, and especially when there is significant lag between patient encounter and survey return. On the flagged surveys, which should be a small fraction of the total, practices could review the free text comments. If the patient comments suggest a perceived diagnostic breakdown, the practice could conduct chart review to determine if a diagnostic error occurred.

Our study has several limitations. In the survey, patients were not specifically asked about diagnostic breakdowns, and patients only rarely used language that explicitly identified a breakdown in diagnosis. Therefore, coders were required to interpret the free text comments, which introduces the possibility of misclassification bias. This limitation was mitigated by extensive coder training and frequent discussion among the research team to reach consensus on interpretation. Second, when we identified that a patient-perceived breakdown occurred, we did not conduct chart review to confirm whether a diagnostic error occurred. Our focus was on the patient's perception of a diagnostic breakdown, which may or may not have been an error. However, we believe that a patient perceiving a breakdown in the diagnostic process is valuable patient safety feedback on its own. Another limitation is that the study did not include survey responses from urgent care and emergency department visits, so the results may not be representative of those settings. The data for our study were collected during the early days of the COVID-19 pandemic. Some free text comments reflect patients' frustration with obtaining ambulatory appointments during

the public health emergency (PHE), while others suggest patients had lower expectations for the care experience during the PHE and were simply grateful to obtain care. Overall, we believe that the PHE may have affected the substance of specific comments but that the study conclusions are still valid. In addition, because of the volume of free text comments, we excluded those comments sentiment coded by the vendor as positive. It is possible that some comments sentiment coded as positive contained diagnostic breakdowns, leading to an underestimation of diagnostic breakdowns, but we expect this to be minimal. Finally, the study was a cross sectional study in a single health system, potentially limiting generalizability. Even with these limitations, the study provides a valuable addition to the literature on using patient feedback to assess diagnostic safety.

CONCLUSIONS

Using patient and family feedback in readily available PREMs is a practical strategy for diagnostic safety measurement in ambulatory care settings. Health systems or individual practices could use the easily monitored structured data as a screening method to identify encounters that may have included a patient-perceived diagnostic breakdown and therefore require further examination. This would allow focusing resources on case review of encounters more likely to have diagnostic safety concerns. Understanding diagnostic breakdowns and measuring diagnostic safety are essential to improving patient safety throughout the diagnostic process, and the patient remains an underutilized resource.

ACKNOWLEDGMENTS

Data availability:

The datasets generated and analyzed during the current study are not publicly available.

REFERENCES

1. The National Academies of Sciences, Engineering, and Medicine. Improving Diagnosis in Health Care. Washington, DC: National Academies Press; 2015:21794.

2. Singh H, Meyer AND, Thomas EJ. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ Qual Saf. 2014;23:727–731.

3. Graber ML, Wachter RM, Cassel CK. Bringing diagnosis into the quality and safety equations. JAMA. 2012;308:1211–1212. [PubMed: 23011708]

4. Agency for Healthcare Research and Quality. Measure Dx: A Resource to Identify, Analyze, and Learn from Diagnostic Safety Events. Rockville, MD: Agency for Healthcare Research and Quality; 2022.

5. Agoritsas T, Bovier PA, Perneger TV. Patient reports of undesirable events during hospitalization. J Gen Intern Med. 2005;20:922–928. [PubMed: 16191139]

6. Iedema R, Allen S, Britton K, et al. What do patients and relatives know about problems and failures in care? BMJ Qual Saf. 2012;21:198–205.

7. Giardina TD, Haskell H, Menon S, et al. Learning from patients' experiences related to diagnostic errors is essential for progress in patient safety. Health Aff (Millwood). 2018;37:1821–1827. [PubMed: 30395513]

8. Aoki T, Watanuki S. Multimorbidity and patient-reported diagnostic errors in the primary care setting: multicentre cross-sectional study in Japan. BMJ Open. 2020;10:e039040.

9. Bontempo AC, Mikesell L. Patient perceptions of misdiagnosis of endometriosis: results from an online national survey. Diagnosis (Berl). 2020;7:97–106. [PubMed: 32007945]

10. Scott J, Heavey E, Waring J, et al. Implementing a survey for patients to provide safety experience feedback following a care transition: a feasibility study. BMC Health Serv Res. 2019;19:613. [PubMed: 31470853]

11. Khan A, Furtak SL, Melvin P, et al. Parent-reported errors and adverse events in hospitalized children. JAMA Pediatr. 2016;170:e154608. [PubMed: 26928413]

12. Khan A, Coffey M, Litterer KP, et al. Families as partners in hospital error and adverse event surveillance. JAMA Pediatr. 2017;171:372–381. [PubMed: 28241211]

13. Mazor KM, Roblin DW, Greene SM, et al. Toward patient-centered cancer care: patient perceptions of problematic events, impact, and response. J Clin Oncol. 2012;30:1784–1790. [PubMed: 22508828]

14. Weissman JS, Schneider EC, Weingart SN, et al. Comparing patient-reported hospital adverse events with medical record review: do patients know something that hospitals do not? Ann Intern Med. 2008;149:100–108. [PubMed: 18626049]

15. Levtzion-Korach O, Frankel A, Alcalai H, et al. Integrating incident data from five reporting systems to assess patient safety: making sense of the elephant. Jt Comm J Qual Patient Saf. 2010;36:402–410. [PubMed: 20873673]

16. O'Hara JK, Reynolds C, Moore S, et al. What can patients tell us about the quality and safety of hospital care? Findings from a UK multicentre survey study. BMJ Qual Saf. 2018;27:673–682.

17. Bell SK, Bourgeois F, Dong J, et al. Patient identification of diagnostic safety blindspots and participation in "good catches" through shared visit notes. Milbank Q. 2022;100:1121–1165. [PubMed: 36539389]

18. Singh H, Bradford A, Goeschel C. Operational measurement of diagnostic safety: state of the science. Diagnosi. 2021;8:51–65.

19. Gurley KL, Wolfe RE, Burstein JL, et al. Use of physician concerns and patient complaints as quality assurance markers in emergency medicine. West J Emerg Med. 2016;17:749–755. [PubMed: 27833684]

20. Gallagher TH, Mazor KM. Taking complaints seriously: using the patient safety lens. BMJ Qual Saf. 2015;24:352–355.

21. Gillespie A, Reader TW. Patient-centered insights: using health care complaints to reveal hot spots and blind spots in quality and safety. Milbank Q. 2018;96:530–567. [PubMed: 30203606]

22. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. BMJ Qual Saf. 2014;23:678–689.

23. Giardina TD, Korukonda S, Shahid U, et al. Use of patient complaints to identify diagnosis-related safety concerns: a mixed-method evaluation. BMJ Qual Saf. 2021;30:996–1001.

24. Agency for Healthcare Research and Quality. CAHPS Clinician & Group Survey. Consumer Assessment of Healthcare Providers and Systems (CAHPS). Published November 2022. Available at: https://www.ahrq.gov/cahps/surveys-guidance/cg/index.html. Accessed February 12, 2024.

25. Nawab K, Ramsey G, Schreiber R. Natural language processing to extract meaningful information from patient experience feedback. Appl Clin Inform. 2020;11:242–252. [PubMed: 32236917]

26. Vandenbroucke JP, Von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. PLoS Med. 2007;4:e297. [PubMed: 17941715]

27. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res. 2005;15:1277–1288. [PubMed: 16204405]

28. Quarfoot D, Levine RA. How robust are multirater interrater reliability indices to changes in frequency distribution? Am Stat. 2016;70:373–384.

29. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–163. [PubMed: 27330520]

30. Harris PA, Taylor R, Thielke R, et al. Research Electronic Data Capture (REDCap) - a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–381. [PubMed: 18929686]

31. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform. 2019;95:103208. [PubMed: 31078660]

32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57:289–300.

33. Sacco AY, Self QR, Worswick EL, et al. Patients' perspectives of diagnostic error: a qualitative study. J Patient Saf. 2021;17:e1759–e1764. [PubMed: 32168272]

34. Street RL Jr., Petrocelli JV, Amroze A, et al. How communication "failed" or "saved the day": counterfactual accounts of medical errors. J Patient Exp. 2020;7:1247–1254. [PubMed: 33457572]

35. Amelung D, Whitaker KL, Lennard D, et al. Influence of doctor-patient conversations on behaviours of patients presenting to primary care with new or persistent symptoms: a video observation study. BMJ Qual Saf. 2020;29:198–208.

36. Bell SK, Bourgeois F, DesRoches CM, et al. Filling a gap in safety metrics: development of a patient-centred framework to identify and categorise patient-reported breakdowns related to the diagnostic process in ambulatory care. BMJ Qual Saf. 2022;31:526–540.

37. Howard M, Fleming ML, Parker E. Patients do not always complain when they are dissatisfied: implications for service quality and patient safety. J Patient Saf. 2013;9:224–231. [PubMed: 24257066]

11,449 unique patients

3,607 coded negative

7,842 coded positive

2,525 confirmed negative

956 recoded as positive or neutral

17 removed for no office visit

109 removed for child visit

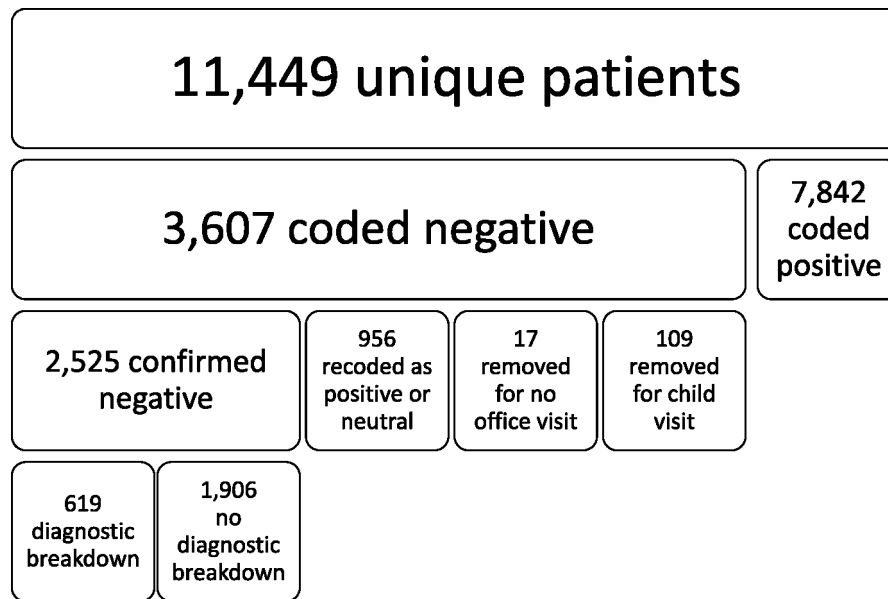619 diagnostic breakdown

1,906 no diagnostic breakdown

FIGURE 1.
Summary of study dataset.

TABLE 1.

Sample Characteristics and Associations With Patient-Perceived Diagnostic Breakdowns

Patient, Appointment, and Survey Characteristics	All Negative Commenters N = 2525	Diagnostic Breakdown n = 619	No Diagnostic Breakdown n = 1906	P
Continuous Characteristics	**Mean (SD)**	**Mean (SD)**	**Mean (SD)**	
Patient age	62.3 (14.3)	61.2 (15.0)	62.6 (14.0)	0.037
Elapsed days (from appointment to survey response)	15.8 (17.7)	17.7 (18.9)	15.2 (17.2)	0.002 [*]
Categorical characteristics	**n (%)**	**n (%)**	**n (%)**	
Patient sex				0.033
Male	881 (34.9)	194 (22.0)	687 (78.2)	
Female	1644 (65.1)	425 (25.9)	1219 (74.2)	
Patient race[†]				0.48
White	1588 (68.5)	378 (23.8)	1210 (76.2)	
Black	598 (25.8)	156 (26.1)	442 (73.9)	
Other	133 (5.7)	35 (26.3)	98 (73.7)	
Visit type[‡]				0.153
In person	931 (39.5)	242 (26.0)	689 (74.0)	
Telehealth	1427 (60.5)	334 (23.4)	1093 (76.6)	
Survey distribution				0.001 [*]
Internet	1928 (76.4)	443 (23.0)	1485 (77.0)	
Paper	597 (23.6)	176 (29.5)	421 (70.5)	

[*]
Significant after Benjamini-Hochberg correction to control false discovery rate.

[†]
Missing data: n = 206 (8.1%).

[‡]
Missing data: n = 167 (6.6%).

TABLE 2.

Associations Between Survey Responses and Patient-Perceived Diagnostic Breakdowns

Questions	Full Sample (N = 2525)	Diagnostic Breakdown	No Diagnostic Breakdown	
Access Domain	**n(%)**	**n (%)**	**n (%)**	***P***
Scheduling appointments				
Low (0–50)	541 (22.0)	155 (26.0)	386 (20.7)	0.006 *
High (51–100)	1921 (78.0)	441 (74.0)	1480 (79.3)	
Receiving return phone calls				
Low (0–50)	637 (27.8)	222 (39.8)	415 (23.9)	<0.001 *
High (51–100)	1655 (72.2)	336 (60.2)	1319 (76.1)	
Contacting practice				
Low (0–50)	852 (35.4)	233 (40.2)	619 (33.9)	0.006 *
High (51–100)	1555 (64.6)	347 (59.8)	1208 (66.1)	
Care provider domain				
Provider show of concern				
Low (0–50)	253 (10.4)	171 (29.3)	82 (4.4)	<0.001 *
High (51–100)	2184 (89.6)	413 (70.7)	1771 (95.6)	
Provider explanations				
Low (0–50)	287 (11.9)	195 (33.6)	92 (5.0)	<0.001 *
High (51–100)	2131 (88.1)	385 (66.4)	1746 (95.0)	
Provider shared decision making				
Low (0–50)	272 (11.4)	181 (32.0)	91 (5.0)	<0.001 *
High (51–100)	2,113 (88.6)	385 (68.0)	1,728 (95.0)	
Provider treatment discussion				
Low (0–50)	305 (12.9)	204 (35.8)	101 (5.6)	<0.001 *
High (51–100)	2064 (87.1)	366 (64.2)	1698 (94.4)	
Provider listening skills				
Low (0–50)	269 (11.1)	188 (32.5)	81 (4.4)	<0.001 *
High (51–100)	2155 (88.9)	391 (67.5)	1764 (95.6)	
Recommendation of provider				

Questions	Full Sample (N = 2525)	Diagnostic Breakdown	No Diagnostic Breakdown	
Access Domain	n(%)	n (%)	n (%)	P
Low (0–50)	348 (14.4)	222 (38.2)	126 (6.9)	<0.001 *
High (51–100)	2065 (85.6)	359 (61.8)	1706 (93.1)	
Personal issues domain				
Treatment with respect/compassion				
Low (0–50)	245 (10.8)	119 (21.7)	126 (7.3)	<0.001 *
High (51–100)	2035 (89.3)	429 (78.3)	1606 (92.7)	
Practice-related domain				
Staff teamwork				
Low (0–50)	356 (15.8)	174 (32.3)	182 (10.7)	<0.001 *
High (51–100)	1892 (84.2)	365 (67.7)	1527 (89.4)	
Recommendation of practice				
Low (0–50)	579 (24.6)	253 (44.4)	326 (18.3)	<0.001 *
High (51–100)	1774 (75.4)	317 (55.6)	1457 (81.7)	

*
Significant after Benjamini-Hochberg correction to control false discovery rate.

TABLE 3.

Effect of Survey Responses on Any Patient-Perceived Diagnostic Breakdowns, Unadjusted and Adjusted for Patient, Appointment, and Survey Characteristics

| | **Any Breakdown** | | |
	Unadjusted Odds Ratio	Adjusted Odds Ratio	95% Confidence Interval
Average access score			
Greater than 50	Reference	Reference	
50 or less	1.14	1.15	0.88–1.51
Average provider score			
Greater than 50	Reference	Reference	
50 or less	6.45	6.82	4.66–9.99 *
Average personal score			
Greater than 50	Reference	Reference	
50 or less	1.19	1.24	0.82–1.89
Average practice score			
Greater than 50	Reference	Reference	
50 or less	2.21	2.33	1.74–3.11 *
Patient age		1.002	0.99–1.01
Elapsed days (from appointment to survey response)		1.01	1.002–1.01 *
Patient sex			
Male		Reference	
Female		1.29	1.04–1.61 *
Visit type			
In-person		Reference	
Telehealth		0.91	0.72–1.15
Survey distribution			
Internet		Reference	
Paper		1.28	0.98–1.7

*
Significant at *P* 0.05.

TABLE 4.

Effect of Survey Responses on Patient-Perceived Diagnostic Breakdowns, by Breakdown Type, Unadjusted and Adjusted for Patient, Appointment, and Survey Characteristics

	Accuracy Breakdown			Timeliness Breakdown			Communication Breakdown		
	Unadjusted Odds Ratio	Adjusted Odds Ratio	95% Confidence Interval	Unadjusted Odds Ratio	Adjusted Odds Ratio	95% Confidence Interval	Unadjusted Odds Ratio	Adjusted Odds Ratio	95% Confidence Interval
Average access score									
Greater than 50	Reference	Reference		Reference	Reference		Reference	Reference	
50 or less	0.66	0.66	0.46–0.97*	2.17	2.26	1.62–3.16*	1.14	1.27	0.90–1.79
Average provider score									
Greater than 50	Reference	Reference		Reference	Reference		Reference	Reference	
50 or less	4.36	5.13	3.42–7.69*	1.47	1.34	0.84–2.14	6.66	6.47	4.39–9.55*
Average personal score									
Greater than 50	Reference	Reference		Reference	Reference		Reference	Reference	
50 or less	1.04	1.05	0.65–1.71	1.55	1.66	1.03–2.69*	1.27	1.33	0.84–2.13
Average practice score									
Greater than 50	Reference	Reference		Reference	Reference		Reference	Reference	
50 or less	2.31	2.08	1.45–2.98*	1.67	1.86	1.25–2.77*	2.44	2.58	1.79–3.71*
Patient age		1.00	0.99–1.01		1.00	0.99–1.01		1.00	0.99–1.01
Elapsed days (from appointment to survey response)		1.01	1.003–1.02*		1.01	1.00–1.02*		1.01	1.00–1.02
Patient sex									
Male		Reference			Reference			Reference	
Female		1.02	0.77–1.35		1.46	1.06–2.02*		1.40	1.03–1.91*
Visit type									
In-person		Reference			Reference			Reference	
Telehealth		1.48	1.07–2.04*		0.70	0.50–0.98*		0.59	0.43–0.82*
Survey distribution									
Internet		Reference			Reference			Reference	
Paper		1.13	0.79–1.60		1.23	0.85–1.78		1.13	0.79–1.63

*
Significant at P 0.05.