Beyond Top-Box Scores: Using Natural Language Processing of Press Ganey Survey Data to Differentiate Enterprise-Wide from Site-Specific Quality Improvement Targets

MAYO CLINIC

Austin Fullenkamp MD, Curtis Simmons MD MBA, Andrew Reuter MD, Easton Neitzel MD, Aileen Park MD MBA, Richard Sharpe MD MBA, Jacob Varner, Nelly Tan MD, Laura Harper MD

Department of Radiology, Mayo Clinic, Phoenix, AZ

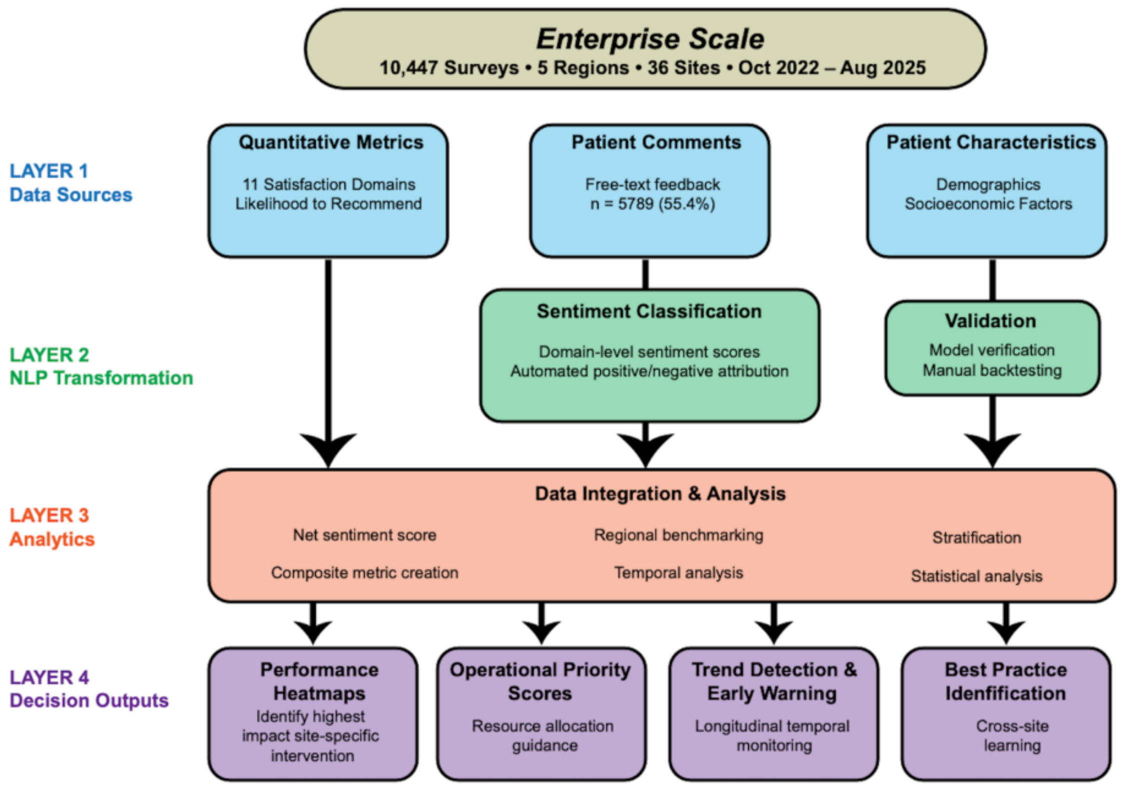
BACKGROUND

Patient experience is often viewed through the lens of top box scores (5/5). However, Press Ganey surveys suffer from ceiling effects eliminating the ability to identify important areas for improvement. We evaluated whether natural language processing (NLP) of textual comments in Press Ganey surveys across a multi-site enterprise could identify operational patterns not seen in aggregated quantitative scores.

METHODS

We analyzed 10,447 surveys from 36 imaging centers in 5 regions (Oct 2022–Aug 2025). Quantitative ratings were uniformly high (enterprise mean 4.81/5.0; regional range 4.72–4.88). Using NLP sentiment classification, we evaluated 5,789 free-text comments across 8 domains (Staff Behavior, Wait Time, Check-in, Physical Environment, Communication, Comfort, Results Delivery, Scheduling) (Figure 1). We calculated net sentiment (% positive minus % negative mentions) by site/region and used ANOVA to test performance gaps. Findings were categorized as universal issues, localized underperformance, or best-practice leaders.

FIGURE 1. MULTI-SITE RADIOLOGY PATIENT EXPERIENCE ANALYSIS



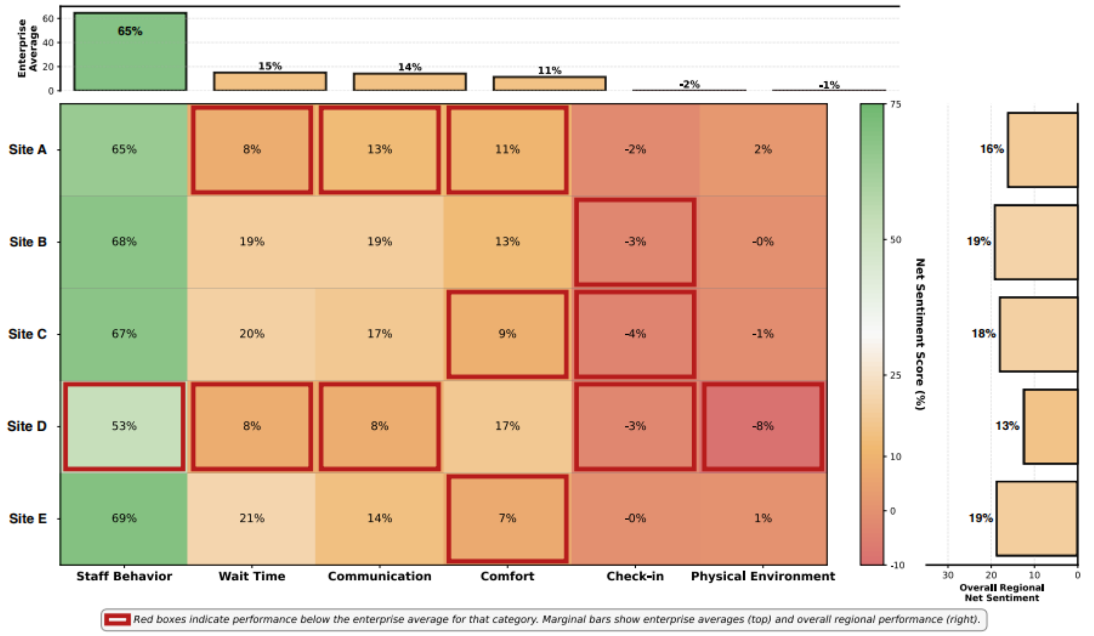
RESULTS

Despite comparable top-box rates (78–91% scoring 5/5), NLP separated **three patterns** (Figure 2):

- Check-in had negative sentiment across all regions (net −2.4% to −0.7%), consistent with an enterprise-wide process problem.
- Wait Time negativity concentrated in a two regions (Site A and D) (net +7.9% vs +20.6% top performer; $p< 0.001$), indicating a targeted operational opportunity.
- Staff Behavior was strongly positive overall (+64.5% net; 4,076 positive mentions), with one region (Site E) significantly outperforming peers (+69.4% vs +65.2% next-best), enabling internal best-practice identification.

Automated processing required < 5 minutes versus an estimated 40+ hours for manual review.

FIGURE 2: REGIONAL NLP SENTIMENT PERFORMANCE MATRIX NET SENTIMENT SCORE (POSITIVE % - NEGATIVE %) WITH ENTERPRISE BENCHMARKS



DISCUSSION

- NLP enables rapid analysis of large volumes of unstructured patient feedback, capturing actionable insights not accessible through quantitative scores alone.
- In this study, NLP differentiated site-specific issues, enterprise-wide challenges, and high-performing regions, all of which were masked by uniformly high top-box scores.
- These findings support prior work demonstrating that narrative analysis can inform targeted operational improvements and enhance patient experience.
- Consistent with existing radiology literature, patient concerns were primarily driven by operational and interpersonal factors rather than diagnostic quality.
- This framework provides a scalable approach to translating patient comments into actionable intelligence at both the site and enterprise level.
- A key limitation is that ~45% of surveys lacked free-text comments, introducing potential response bias.
- Future work will focus on implementing targeted interventions based on NLP findings and assessing their impact on patient experience metrics.

CONCLUSIONS

Press Ganey top-box scores showing similar satisfaction across all enterprise regions provides no actionable differentiation for quality improvement prioritization. When top-box scores cannot differentiate priorities, NLP converts Press Ganey narratives into scalable operational intelligence supporting enterprise-wide fixes, targeted interventions, and spread of internal best practices. This study provides a practical framework for radiology business analytics, demonstrating how informatics tools reveal improvement priorities hidden within apparently uniform satisfaction data.

REFERENCES

1. Khanbhai M, Anyadi P, Symons J, et al. Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. BMJ Health Care Inform. 2021;28(1):e100262. doi:10.1136/bmjhci-2020-100262
2. Nawab K, Ramsey G, Schreiber R. Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback. Appl Clin Inform. 2020;11(2):242–252. doi:10.1055/s-0040-1708049
3. Nguyen E, Dodoo CA, Banerjee I, et al. Effects of patient survey feedback on improving patient experience with outpatient magnetic resonance imaging: Patient Feedback Helps Improve MRI Experience. Curr Probl Diagn Radiol. 2025;54(3):369–376. doi:10.1067/j.cpradiol.2024.10.035
4. Terpend R, Rossetti C, Kroes J, Mudge S, Glass J. Leveraging Free-Form Comments to Assess and Improve Patient Satisfaction. Ann Fam Med. 2022;20(6):551–555. doi:10.1370/afm.2888
5. Kleiss IIM, Kortlever JTP, Karyampudi P, et al. A Comparison of 4 Single-Question Measures of Patient Satisfaction. J Clin Outcomes Manag. 2020;27(1):41–47.
6. Lang EV, Yuh WTC, Ajam A, et al. Understanding Patient Satisfaction Ratings for Radiology Services. AJR Am J Roentgenol. 2013;201(6):1190–1196. doi:10.2214/AJR.13.11281