JURNAL LOCUS: Penelitian & Pengabdian
Volume 5 No. 6 June 2026
E-ISSN 2829-7334| P-ISSN 2829-5439
Hompage: https://locus.rivierapublishing.id/index.php/jl

Complaint Severity and Resolution Time in a Hospital Outpatient Department Digital Complaint Management System: A Retrospective Observational Study

Michelly, Mirachel, Puspita Angelina*, Thoe Isabella Roseline Pramana Putri
Universitas Pelita Harapan, Indonesia
Email: michelly.tjoa@gmail.com, mirachel17@gmail.com, puspita.angelina13@gmail.com*, lala.thoe@gmail.com

Keywords:
patient complaint management; complaint severity; turnaround time; outpatient department; healthcare quality; digital complaint system; retrospective study

Abstract

Efficient management of patient complaints is a cornerstone of healthcare quality assurance. This study aimed to determine the association between complaint severity level and overall turnaround time (TAT) in a hospital OPD digital complaint management system. A retrospective observational study was conducted using a balanced sample of 96 OPD complaint tickets (24 per severity group, S1–S4) extracted from a tertiary hospital digital complaint system. The primary outcome was Overall TAT (minutes). Board TAT and complaint type were included as secondary variables. Normality was assessed using the Shapiro–Wilk test. The Kruskal–Wallis H test with Bonferroni post-hoc correction was applied for between-group comparisons. Pearson correlation examined bivariate associations. Multiple linear regression was performed to assess the simultaneous predictive value of severity group, Board TAT, and complaint type on Overall TAT. The sample comprised 58 males (60.4%) and 38 females (39.6%). Overall TAT differed significantly across severity groups ($H = 55.344$, $p < 0.001$), with S1 having the highest median TAT (27.07 min) and S4 the lowest (0.00 min). Board TAT also differed significantly across groups ($H = 58.275$, $p < 0.001$), with significant pairwise differences between S1 vs S2 ($p = 0.010$) and S1 vs S4 ($p = 0.003$). Pearson correlation revealed a significant inverse association between severity group and Overall TAT ($r = -0.215$, $p = 0.018$). The multivariable regression model was not statistically significant ($F(3,92) = 1.747$, $p = 0.163$; $R^2 = 0.054$), suggesting that additional unmeasured factors contribute to resolution time. Complaint severity is significantly associated with Overall TAT at the bivariate level, with higher-urgency complaints demonstrating longer resolution times.

INTRODUCTION

Patient complaints are a valuable source of information about care quality, often revealing deficiencies or system failures that formal audits can overlook (Rhys et al., 2024). When addressed promptly and openly, such feedback can strengthen patient trust and drive institutional learning; conversely, complaints that are mishandled or delayed can undermine confidence in care and even expose the hospital to legal risk (Khaled et al., 2025). On average, managing complaints demands substantial effort – one study estimated over 3,300 person-hours per year on complaint investigation in a single large hospital (Lee et al., 2018) – yet patient dissatisfaction with complaint handling remains high. This suggests that complaint management is not only a legal and ethical obligation but also a critical component of quality assurance and patient-centred care improvement (Rhys et al., 2024; Lee et al., 2018).

In modern hospitals, patient grievances are frequently logged in digital ticketing systems that timestamp each stage of resolution and assign a severity rating. Typical severity scales (e.g. "S1" urgent down to "S4" low-priority) are intended to triage resources and ensure that urgent issues are escalated quickly. However, it is unclear whether the assigned severity level reliably predicts how long a case actually remains open (Moons et al., 2015, 2019). Few studies have examined this directly. Lee et al. (2018) found in a retrospective cohort that complaints deemed more severe required more staff-hours to manage, but this work measured effort rather than elapsed calendar time, and covered a mixed inpatient/outpatient population. To our knowledge, no study has analyzed full complaint-to-closure turnaround times in a single-system digital complaint database, especially focusing on outpatient (OPD) cases.

The outpatient department (OPD) is a particularly important setting because it generates the highest volume of patient contacts in a tertiary hospital, spanning registration, clinical consultation, pharmacy, and billing. Common themes in OPD complaints include long waiting times, poor communication or information provision, staff attitudes, and billing disputes (Yong et al., 2024; Li et al., 2024). Importantly, each of these issues maps to distinct process stages (e.g. front-desk wait, clinic interaction, cashier/billing) with measurable turnaround times. Understanding which features of OPD complaints (such as severity, category, or initial response lag) predict longer resolution times could therefore inform targeted improvements (Ahsan et al., 2023; Aris-Brosou et al., 2018; Giardina et al., 2021).

Despite this potential, most published work on healthcare complaints has been descriptive – reporting incidence, categories or patient satisfaction – without quantitatively linking complaint attributes to resolution timelines (Bayer et al., 2021; Elias et al., 2021; Hansen et al., 2025; Van Dael et al., 2020). Studies that do examine resolution have typically measured staff effort or looked at mixed populations (Lee et al., 2018), rather than elapsed days to closure in a pure OPD context, rather than elapsed days to closure in a pure OPD context. No prior research, to our knowledge, has used the rich timestamp data from a digital OPD complaint system to trace the full complaint-to-closure trajectory.

The urgency of this research is underscored by several factors. First, patient satisfaction with complaint handling remains low in many healthcare systems, and lengthy resolution times are a frequently cited source of dissatisfaction. Second, healthcare accreditation bodies increasingly require hospitals to demonstrate timely and effective complaint management, making evidence-based performance benchmarks necessary (Araujo et al., 2020; Devkaran & O'Farrell, 2015; Mosadeghrad, 2021). Third, the digitisation of complaint systems in Indonesian hospitals has created new opportunities for data-driven quality improvement, but these data remain largely underutilised for research purposes. Fourth, the growing patient population in Indonesian tertiary hospitals has led to increasing complaint volumes, making efficient complaint resolution a practical necessity rather than merely an academic concern. Therefore, understanding the determinants of complaint resolution time has immediate practical implications for hospital quality management (Aboalganam et al., 2024; August et al., 2025; van Dael et al., 2022).

The novelty of this research lies in several key aspects. First, this is the first study to systematically analyse the relationship between severity classification and actual turnaround time in a digital OPD complaint management system in an Indonesian hospital setting. Second, the study uses rich timestamp data to trace the full complaint-to-closure trajectory, including

both initial response time (Board TAT) and overall resolution time. Third, the research employs a balanced sampling strategy that enables robust statistical comparison across all four severity levels (S1–S4), including the highest-urgency S1 cases that are typically rare in naturalistic samples. Fourth, the study integrates bivariate correlation, non-parametric group comparison, and multivariable regression analysis to provide a comprehensive assessment of the predictors of resolution time. Fifth, the identification of the non-significant multivariable model despite bivariate significance points to the complexity of the complaint resolution process and the likely influence of unmeasured contextual factors, opening new avenues for future research.

The present study addresses this gap by analyzing OPD complaint tickets resolved in the hospital's digital system. Our primary aim is to determine whether the severity level assigned to each complaint predicts overall turnaround time (TAT) from lodging to closure, controlling for factors such as initial board response delay and complaint type. As a secondary aim, we characterize the resolution profiles for each severity class, with particular attention to how the highest-urgency (S1) and lowest-urgency (S4) cases differ in their time-to-close profiles. This analysis provides data-driven insight into complaint management performance and may suggest refinements to triage protocols and service level agreements.

METHODS

Study Design and Setting

This was a retrospective observational study using administrative data from the digital complaint management system of a single tertiary hospital. Data were extracted from the OPD form-type category of the system's complaint ticketing platform, which records patient feedback across multiple submission channels (direct, telephone, and digital messaging). The study was conducted in accordance with the Declaration of Helsinki. Ethics committee approval or waiver was obtained in accordance with institutional policy.

Study Population and Sampling

The target population comprised all OPD complaint tickets registered in the hospital digital system during the study period. To enable between-group statistical comparison across four severity levels with adequate statistical power, a balanced sampling strategy was applied: 24 tickets were selected from each severity group (S1–S4), yielding a total analytic sample of n = 96. All S1 OPD tickets available in the dataset were included (n = 24, representing the total S1 OPD population). For S2, S3, and S4, random sampling was applied to select 24 tickets per group from the available records.

Variable Definitions

The primary outcome was *Overall TAT*, defined as the total elapsed time in minutes from complaint ticket creation (Ticket Board Open On) to final closure (Ticket Board Close On). The primary predictor was *Severity group* (ordinal scale: S1 = most urgent to S4 = lowest priority), assigned by complaint officers at ticket intake. *Board TAT* (minutes from ticket creation to first board-level response) was included as a process-level covariate. *Complaint type* was a derived categorical variable grouping raw complaint subcategories into five thematic domains: Waiting time, Information/Guidance, Staff behaviour, Billing/Pricing, and Others. Patient *Gender* (Male/Female) was reported descriptively.

Statistical Analysis

Data were analysed using IBM SPSS Statistics version 27 (IBM Corp., Armonk, NY, USA). Normality of Overall TAT and Board TAT within each severity group was assessed using the Shapiro–Wilk test. Given that both variables violated the normality assumption across all groups (all $p < 0.05$), non-parametric methods were used for group comparisons. The Kruskal–Wallis H test was applied to compare Overall TAT and Board TAT across the four severity groups. Where significant overall differences were detected, pairwise post-hoc comparisons were performed using the Bonferroni correction. Pearson correlation coefficients were calculated to examine bivariate associations between Overall TAT and each of the three candidate predictors (severity group, Board TAT, complaint type). Multiple linear regression was then performed with Overall TAT as the dependent variable and severity group, Board TAT, and complaint type as independent variables, entered simultaneously (Enter method). Model fit was evaluated using the F-statistic, R^2, and Adjusted R^2. Standardised (β) and unstandardised (B) coefficients with 95% confidence intervals (CI) are reported. Statistical significance was set at $\alpha = 0.05$ (two-tailed).

Hypothesis

H0 (Null Hypothesis): There is no relationship between complaint severity and Overall TAT.
H1 (Alternative Hypothesis): There is a relationship between complaint severity and Overall TAT.

RESULT AND DISCUSSION

Sample Characteristics

A total of 96 OPD complaint tickets were included in the analysis, with 24 cases per severity group (S1–S4). Of the 96 participants, 58 (60.4%) were male and 38 (39.6%) were female. The most common complaint type was Waiting time (n = 30; 31.3%), followed by Others (n = 23; 24.0%), Information/Guidance (n = 19; 19.8%), Staff behaviour (n = 16; 16.7%), and Billing/Pricing (n = 8; 8.3%). Full sample characteristics are presented in Table 1.

Table 1. Characteristics of respondents based on severity level, gender, and complaint type (n = 96)

Variable	Classification	n (%)
Severity level	Severity 1 (S1)	24 (25.0%)
	Severity 2 (S2)	24 (25.0%)
	Severity 3 (S3)	24 (25.0%)
	Severity 4 (S4)	24 (25.0%)
Gender	Male	58 (60.4%)
	Female	38 (39.6%)
Complaint type	Waiting time	30 (31.3%)
	Information/Guidance	19 (19.8%)
	Staff behaviour	16 (16.7%)
	Billing/Pricing	8 (8.3%)
	Others	23 (24.0%)

Normality Assessment

The Shapiro–Wilk test indicated that both Overall TAT and Board TAT were non-normally distributed within each severity group (all $p < 0.05$), except for Severity 4, in which TAT values were invariant (all observations = 0.00, consistent with near-instantaneous administrative closure). Non-parametric tests were therefore applied for all between-group comparisons.

Kruskal–Wallis Comparisons by Severity Group

Overall TAT differed significantly across the four severity groups (Kruskal–Wallis $H = 55.344$, $df = 3$, $p < 0.001$). S1 complaints had the highest median Overall TAT (27.07 min; range 0.00– 28,882.78 min), followed by S2 (17.32 min; 0.12–67.98 min) and S3 (2.19 min; 0.02–26.65 min), whereas S4 had a median Overall TAT of 0.00 min (range 0.00–0.00 min), consistent with its confirmed identity as a low-priority administrative suggestion pathway. Notwithstanding the significant overall H statistic, Bonferroni-corrected pairwise post-hoc comparisons did not identify any statistically significant differences between individual pairs of severity groups for Overall TAT (all adjusted $p > 0.05$; see Table 3).

Board TAT also differed significantly across severity groups ($H = 58.275$, $df = 3$, $p < 0.001$). S1 had the highest median Board TAT (2.62 min; 0.02–118.67 min), followed by S3 (1.85 min; 0.02–15.17 min) and S2 (0.84 min; 0.12–7.35 min), with S4 again at 0.00 min. Post-hoc Bonferroni comparisons revealed significant pairwise differences between S1 and S2 ($p = 0.010$) and between S1 and S4 ($p = 0.003$; see Table 4).

Table 2. Overall TAT and Board TAT by severity group (n = 96)

Variable	Severity Group	Median (min–max), minutes	p-value[a]
Overall TAT	Severity 1 (S1)	27.07 (0.00–28,882.78)*	< 0.001[b]
	Severity 2 (S2)	17.32 (0.12–67.98)*	
	Severity 3 (S3)	2.19 (0.02–26.65)*	
	Severity 4 (S4)	0.00 (0.00–0.00)	
Board TAT	Severity 1 (S1)	2.62 (0.02–118.67)*	< 0.001[b]
	Severity 2 (S2)	0.84 (0.12–7.35)*	
	Severity 3 (S3)	1.85 (0.02–15.17)*	
	Severity 4 (S4)	0.00 (0.00–0.00)	

[a] Kruskal–Wallis H test. [b] Statistically significant ($p < 0.05$). * Non-normal distribution (Shapiro–Wilk $p < 0.05$). TAT = Turnaround Time.

Table 3. Bonferroni post-hoc pairwise comparisons: Overall TAT between severity groups

Groups	S1	S2	S3	S4
Severity 1 (S1)		NS (p = 0.160)	NS (p = 0.159)	NS (p = 0.156)
Severity 2 (S2)	NS (p = 0.160)		NS (p = 1.000)	NS (p = 1.000)
Severity 3 (S3)	NS (p = 0.159)	NS (p = 1.000)		NS (p = 1.000)
Severity 4 (S4)	NS (p = 0.156)	NS (p = 1.000)	NS (p = 1.000)	

NS = Not Significant. Bonferroni correction applied. No pairwise comparison reached statistical significance for Overall TAT despite the significant Kruskal–Wallis H statistic,

likely reflecting inadequate power for pairwise comparisons and the large variance in the S1 group.

Table 4. Bonferroni post-hoc pairwise comparisons: Board TAT between severity groups

Groups	S1	S2	S3	S4
Severity 1 (S1)		S (p = 0.010)	NS (p = 0.081)	S (p = 0.003)
Severity 2 (S2)	S (p = 0.010)		NS (p = 1.000)	NS (p = 1.000)
Severity 3 (S3)	NS (p = 0.081)	NS (p = 1.000)		NS (p = 1.000)
Severity 4 (S4)	S (p = 0.003)	NS (p = 1.000)	NS (p = 1.000)	

S = Significant; NS = Not Significant. Bold cells indicate statistically significant pairwise differences (Bonferroni-corrected $p < 0.05$). Significant pairs: S1 vs S2 ($p = 0.010$); S1 vs S4 ($p = 0.003$). TAT = Turnaround Time.

Pearson Correlation Analysis

Pearson correlation analysis revealed a statistically significant inverse association between severity group and Overall TAT ($r = -0.215$, $p = 0.018$), indicating that higher-urgency complaints (lower severity number) were associated with longer resolution times. Board TAT showed a positive but non-significant correlation with Overall TAT ($r = 0.143$, $p = 0.082$). Complaint type was not significantly correlated with Overall TAT ($r = -0.074$, $p = 0.237$). Correlation statistics are summarised in Table 5.

Table 5. Pearson correlation coefficients between Overall TAT and candidate predictors

Variable	r	p-value
Severity group vs Overall TAT	−0.215*	0.018*
Board TAT vs Overall TAT	0.143	0.082
Complaint type vs Overall TAT	−0.074	0.237

* Statistically significant at $p < 0.05$ (two-tailed). Pearson correlation; $n = 96$. TAT = Turnaround Time.

Multiple Linear Regression Analysis

A multiple linear regression analysis was performed to evaluate the simultaneous predictive value of severity group, Board TAT, and complaint type on Overall TAT. Preliminary bivariate correlation analyses indicated that only severity group shared a statistically significant negative correlation with Overall TAT ($r = -0.215$, $p = 0.018$), whereas Board TAT ($r = 0.143$, $p = 0.082$) and complaint type ($r = -0.074$, $p = 0.237$) did not reach statistical significance at the bivariate level.

When entered simultaneously into the regression model, the combination of the three predictors did not significantly predict Overall TAT, $F(3, 92) = 1.747$, $p = 0.163$. The model accounted for a small proportion of total variance, with $R^2 = 0.054$ (Adjusted $R^2 = 0.023$), indicating that the three included predictors together explained only 5.4% of the variability in Overall TAT.

As detailed in Table 6, none of the individual predictors achieved statistical significance within the multivariable model. Severity group, despite its bivariate significance, was not an

independent predictor within the multivariable model (β = −0.208, B = −588.933, p = 0.092; 95% CI [−1275.454, 97.587]). Similarly, Board TAT (β = 0.083, p = 0.438) and complaint type (β = 0.037, p = 0.750) demonstrated no significant independent association with Overall TAT in the simultaneous model.

Table 6. Multiple linear regression: predictors of Overall TAT (n = 96)

Variable	B	SE	β	t	p	95% CI
Severity group	−588.933	345.665	−0.208	−1.704	0.092	[−1275.454, 97.587]
Board TAT	19.680	25.267	0.083	0.779	0.438	[−30.502, 69.862]
Complaint type	75.785	237.606	0.037	0.319	0.750	[−396.122, 547.692]
(Constant)	1685.592	887.919		1.898	0.061	[−77.892, 3449.077]

Model: $R^2 = 0.054$; Adjusted $R^2 = 0.023$; $F(3, 92) = 1.747$; $p = 0.163$ (not significant). Dependent variable: Overall TAT (minutes). B = Unstandardised coefficient; SE = Standard Error; β = Standardised coefficient; CI = Confidence Interval.

Principal Findings

This retrospective study of 96 balanced OPD complaint tickets from a tertiary hospital digital complaint management system yields three principal findings. First, complaint severity is significantly associated with Overall TAT at the bivariate level: the Kruskal–Wallis test confirmed highly significant differences in resolution time across severity groups ($H = 55.344$, $p < 0.001$), with S1 complaints requiring the longest median resolution time (27.07 min) and S4 the shortest (0.00 min). Second, Pearson correlation confirmed a significant inverse association between severity group and Overall TAT ($r = −0.215$, $p = 0.018$), indicating that higher-urgency complaints are resolved more slowly despite being assigned greater priority. Third, the multivariable linear regression model incorporating severity group, Board TAT, and complaint type did not reach statistical significance ($R^2 = 0.054$, $p = 0.163$), suggesting that these three variables together explain only a small fraction of the total variability in Overall TAT.

Interpretation of Bivariate Findings

The significant inverse correlation between severity group and Overall TAT is clinically meaningful and directionally consistent with expectations: higher-urgency complaints (lower severity number) take longer to resolve in absolute terms. This finding mirrors prior evidence from Lee et al. (2018), who demonstrated that more severe complaint domains required substantially greater person-hours to manage, even when controlling for complaint volume. In the current study, S1 complaints had a median Overall TAT of 27.07 minutes with an extreme maximum of 28,882.78 minutes, reflecting substantial heterogeneity in resolution complexity even within the most urgent category. This wide variance likely reflects the diversity of S1 complaint types ranging from clinical care quality concerns to patient safety incidents each with markedly different investigation requirements.

The significant Board TAT difference between S1 and S2 ($p = 0.010$) and between S1 and S4 ($p = 0.003$) in the post-hoc analysis indicates that the digital complaint system does differentiate its initial response speed based on severity S1 complaints receive faster board-level attention relative to the adjacent group (S2). The non-significant difference between S1

and S3 ($p = 0.081$) in Board TAT, however, suggests that the prioritisation gradient is not perfectly monotonic across the full severity scale, pointing to possible inconsistency in triage practice that warrants operational review.

Interpretation of the Non-Significant Multivariable Model

The non-significance of the multivariable linear regression model ($p = 0.163$) should not be interpreted as an absence of a relationship between severity and resolution time. Rather, it reflects the complexity of the complaint resolution process and points to the influence of unmeasured contextual variables not captured in the current dataset. When severity group, Board TAT, and complaint type are entered simultaneously, each variable's independent contribution becomes attenuated, likely because these predictors are moderately intercorrelated (severity group and complaint type: $r = 0.494$, which is substantial), and because the three-variable model does not capture the full pathway from complaint receipt to closure.

The loss of bivariate significance for severity group in the multivariable model ($\beta = -0.208$, $p = 0.092$) is consistent with partial mediation or confounding: complaint type which is strongly associated with severity group ($r = 0.494$) may capture some of the variance in Overall TAT that was originally attributed to severity alone. This finding suggests that the effect of severity on resolution time is not a simple direct effect but is channelled, in part, through the type of complaint raised. Future research should explore whether formal mediation models, incorporating intermediate variables such as investigation complexity, staff involvement level, or departmental workload, can better account for the variance in overall TAT that the current model leaves unexplained.

Clinical and Operational Implications

Despite the non-significant regression model, the bivariate evidence presented in this study carries actionable implications for hospital quality management. The finding that S1 complaints take significantly longer to resolve than lower-urgency complaints despite presumably receiving priority attention suggests that the current complaint resolution workflow does not provide sufficient structural support for the most complex cases. Service level agreements (SLAs) should be differentiated by severity, with realistic but binding maximum resolution timeframes for S1 cases that account for their inherent complexity rather than applying a uniform TAT target across all severity groups.

The near-zero Overall TAT and Board TAT for S4 complaints is consistent with their operational designation as low-priority general suggestions requiring only administrative acknowledgement rather than substantive investigation. This group should be managed through a separate expedited administrative pathway and should be excluded from standard complaint TAT performance benchmarks to avoid artificially inflating system-level resolution speed metrics.

Limitations

Several limitations of this study warrant acknowledgement. First, this is a single-centre study, and findings may not generalise to hospitals with different complaint management workflows, staffing models, or patient demographics. Second, the balanced sampling strategy while enabling group comparisons overrepresents S1 complaints relative to their natural population prevalence and underrepresents S3 and S4; the sample thus reflects a constructed rather than naturalistic severity distribution. Third, the small sample size within each group ($n = 24$) limits statistical power for pairwise post-hoc comparisons and multivariable

regression, and may account for the non-significant Bonferroni comparisons for Overall TAT despite the significant overall Kruskal–Wallis result. Fourth, the extreme upper range of S1 Overall TAT (28,882.78 minutes) suggests the presence of at least one severe outlier that may disproportionately influence mean estimates and regression coefficients; robust regression methods or sensitivity analyses excluding extreme outliers would strengthen future analyses. Fifth, the three predictors included in the regression model explain only 5.4% of Overall TAT variance, indicating that important determinants such as investigation complexity, escalation status, responsible department, and time of complaint submission were not captured.

CONCLUSION

Complaint severity is significantly associated with overall TAT in a hospital OPD digital complaint management system at the bivariate level, with higher-urgency complaints (S1) demonstrating the longest resolution times and lower-urgency administrative suggestions (S4) being closed near-instantaneously. However, the multivariable regression model incorporating severity group, Board TAT, and complaint type did not reach statistical significance, accounting for only 5.4% of Overall TAT variance. This finding reflects the systemic complexity of complaint resolution and the likely influence of unmeasured process variables. These results support three actionable recommendations for hospital quality management: (1) the development of severity-tiered SLAs with differentiated maximum resolution timeframes for S1–S3 complaints; (2) the separation of S4 administrative suggestions from standard complaint performance metrics; and (3) the incorporation of additional process variables including investigation type, escalation status, and departmental workload into future analyses to better account for the determinants of resolution time. A prospective, multi-centre study with a larger and more balanced sample would strengthen the evidence base for complaint management system design.

DAFTAR PUSTAKA

Aboalganam, K. M., Alzghoul, A., & Alhanatleh, H. (2024). An analysis of service quality and complaint handling in the Jordanian healthcare sector: Implications for TQM and customer retention. *Innovative Marketing*.

Ahsan, F., Dana, N. H., Sarker, S., Li, Muyeen, S., Ali, M. F., Tasneem, Z., Hasan, M., Abhi, S. H., Islam, M. R., Ahamed, M. H., Islam, M. M., Das, S. K., Badal, M. F. R., & Das, P. (2023). Data-driven next-generation smart grid towards sustainable energy evolution: Techniques and technology review. *Protection and Control of Modern Power Systems*.

Araujo, C. A. S., Siqueira, M. M., & Malik, A. M. (2020). Hospital accreditation impact on healthcare quality dimensions: A systematic review. *International Journal for Quality in Health Care*.

Aris-Brosou, S., Kim, J., Li, X., & Liu, H. (2018). Predicting the reasons of customer complaints: A first step toward anticipating quality issues of in vitro diagnostics assays with machine learning. *JMIR Medical Informatics*.

August, F., Nyamhanga, T., Kakoko, D., Anaeli, A., & Mwangu, M. (2025). Implementation of patient complaint management mechanisms for quality improvement in maternal health services: Exploration of provider perspectives at Mkuranga District Hospital. *International Journal of Women's Health*.

Bayer, S., Kuzmickas, P., Boissy, A., Rose, S. L., & Mercer, M. (2021). Categorizing and rating patient complaints: An innovative approach to improve patient experience.

Journal of Patient Experience.
Ching, E. S. Y., Chia, I. W. Y., Low, S. H., Ang, P. Y., Koh, S. W. C., Wong, E. R. N., et al. (2025). Classifying patient complaints using artificial intelligence–powered large language models: Cross-sectional study. *Journal of Medical Internet Research, 27*, e74231. https://doi.org/10.2196/74231
Devkaran, S., & O'Farrell, P. N. (2015). The impact of hospital accreditation on quality measures: An interrupted time series analysis. *BMC Health Services Research*.
Elias, R., Fischer, K. M., Siddiqui, M., Coons, T., Meyerhofer, C. A., Pretzman, H. J., Greig, H. E., Stevens, S., & Burton, M. (2021). A taxonomic review of patient complaints in adult hospital medicine. *Journal of Patient Experience*.
Giardina, T., Korukonda, S., Shahid, U., Vaghani, V., Upadhyay, D. K., Burke, G. F., & Singh, H. (2021). Use of patient complaints to identify diagnosis-related safety concerns: A mixed-method evaluation. *BMJ Quality & Safety*.
Hansen, S. M.-B., Clausen, M. K., Hansen, N., Eriksen, M. B., Sørensen, A., Bogh, S. B., Birkeland, S., & Morsø, L. (2025). From complaint material to quality improvement: Exploring the use of patient complaints or compensation claims in quality improvement initiatives—A scoping review. *PubMed*.
Khaled, A., Sabir, M., Qureshi, R., Caruso, M., Guarrasi, V., Xiang, S., & Zhou, K. (2025). Leveraging MIMIC datasets for better digital health: A review on open problems, progress highlights, and future promises. *arXiv*.
Lee, J., Loh, T. P., Ong, D. E. H., Caleb, M. G., Lim, A. Y. T., & Manning, P. G. (2018). Time needed to resolve patient complaints and factors influencing it: A cohort study. *International Journal for Quality in Health Care, 30*(7), 571–575. https://doi.org/10.1093/intqhc/mzy088
Li, G., Chen, Y., & Lou, X. (2024). Complaint management system and patient satisfaction in grassroots hospitals. *Medicine, 103*(8), e37275. https://doi.org/10.1097/MD.0000000000037275
Moons, K., Altman, D., Reitsma, J., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E., Vickers, A., Ransohoff, D., & Collins, G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. *Annals of Internal Medicine*.
Moons, K., Wolff, R., Riley, R., Whiting, P., Westwood, M., Collins, G., Reitsma, J., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Annals of Internal Medicine*.
Mosadeghrad, A. M. (2021). Hospital accreditation: The good, the bad, and the ugly. *International Journal of Healthcare Management*.
Rhys, C. S., Benwell, B., Erofeeva, M., & Simmons, R. (2024). Harnessing the power of language to enhance patient experience of the NHS complaint journey in Northern Ireland: A mixed-methods study. *Health and Social Care Delivery Research, 12*(33), 1–224. https://doi.org/10.3310/LNYW7741
van Dael, J., Reader, T. W., Gillespie, A., Freise, L., Darzi, A., & Mayer, E. K. (2022). Do national policies for complaint handling in English hospitals support quality improvement? Lessons from a case study. *Journal of the Royal Society of Medicine*.
van Dael, J., Reader, T. W., Gillespie, A., Neves, A. L., Darzi, A., & Mayer, E. (2020). Learning from complaints in healthcare: A realist review of academic literature, policy evidence and front-line insights. *BMJ Quality & Safety*.
Yong, L. P. X., Tung, J. Y. M., Lee, Z. Y., Kuan, W. S., & Chua, M. T. (2024). Performance of large language models in patient complaint resolution: Web-based cross-sectional survey. *Journal of Medical Internet Research, 26*, e56413. https://doi.org/10.2196/56413