



FINANCIAL OUTCOMES OF CLINICAL PRACTICE VARIATION UNDER PROSPECTIVE PAYMENT SYSTEMS: A BILLING-DATA AUDIT

Fallen Wicaksono*, Firman Pribadi

Fakultas Kedokteran dan Ilmu Kesehatan, Universitas Muhammadiyah Yogyakarta, Jl. Lkr. Sel., Geblagan, Tamantirto, Kec. Bantul, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55183, Indonesia

*fallenwicaksono@gmail.com

ABSTRAK

Di bawah sistem pembayaran prospektif INA-CBG, rumah sakit menanggung kerugian ketika biaya perawatan sebenarnya melebihi tarif tetap. Lama rawat inap (LOS) adalah indikator efisiensi yang umum, namun LOS yang sebanding mungkin tidak mencerminkan penggunaan sumber daya yang sebanding. Studi ini mengevaluasi variasi tingkat dokter dalam pengeluaran farmasi dan margin keuangan di antara episode rawat inap yang dikodekan sebagai K-4-17-I. Audit penagihan retrospektif dari 856 episode rawat inap gastroenteritis ringan atau nyeri perut yang dikodekan INA-CBG K-4-17-I selama tahun fiskal 2025 membandingkan LOS, tarif, total biaya sebenarnya, biaya farmasi, dan margin keuangan di tiga kelompok dokter anonim (A, B, C). Variabel miring dilaporkan sebagai median (IQR) dan dibandingkan dengan Kruskal-Wallis dengan pengujian post-hoc Dunn dan koreksi Bonferroni; regresi median dan model linier umum gamma-log disesuaikan untuk usia, jenis kelamin, diagnosis sekunder, dan LOS. Median LOS adalah 3 hari di semua kelompok; Hanya perbandingan B versus C yang lolos uji korelasi (p yang disesuaikan = 0,012, $\epsilon^2 = 0,010$), perbedaan yang secara klinis tidak signifikan. Biaya farmasi dan margin keuangan berbeda secara signifikan (keduanya $p < 0,001$): biaya farmasi median adalah IDR 551.439 (C), 342.097 (A), dan 285.397 (B); margin median IDR 309.097 (B), -2.290 (A), dan -134.296 (C). Setelah penyesuaian, margin median IDR 237.522 dan 394.194 lebih rendah di A dan C daripada di B, dan biaya farmasi 1,18 dan 1,71 kali lebih tinggi (semua $p < 0,001$). Lama rawat inap yang sebanding tidak menjamin konsumsi sumber daya yang seragam; variasi biaya tetap ada setelah penyesuaian untuk campuran kasus yang tercatat dan mungkin tetap tidak terdeteksi ketika efisiensi dipantau hanya berdasarkan durasi rawat inap. Desain retrospektif menghalangi penentuan penyebab. Rumah sakit harus melengkapi pemantauan lama rawat inap (LOS) dengan audit biaya kelompok dokter secara agregat dan tata kelola jalur klinis yang terdigitalisasi.

Kata kunci: audit penagihan rumah sakit; biaya farmasi; INA-CBG; jalur klinis; margin keuangan; variasi praktik klinis

FINANCIAL OUTCOMES OF CLINICAL PRACTICE VARIATION UNDER PROSPECTIVE PAYMENT SYSTEMS: A BILLING-DATA AUDIT

ABSTRACT

Under INA-CBG prospective payment, hospitals absorb losses when real treatment costs exceed fixed tariffs. Length of stay (LOS) is a common efficiency indicator, yet comparable LOS may not reflect comparable resource use. This study evaluated physician-level variation in pharmaceutical expenditure and financial margins among inpatient episodes coded as K-4-17-I. A retrospective billing audit of 856 inpatient episodes of mild gastroenteritis or abdominal pain coded INA-CBG K-4-17-I during the 2025 fiscal year compared LOS, tariff, total real cost, pharmacy cost, and financial margin across three anonymized physician groups (A, B, C). Skewed variables are reported as median (IQR) and compared by Kruskal-Wallis with Dunn post-hoc testing and Bonferroni correction; median regression and gamma-log generalized linear models adjusted for age, sex, secondary diagnosis, and LOS. Median LOS was 3 days in all groups; only the B versus C comparison survived correction (adjusted $p = 0.012$, $\epsilon^2 = 0.010$), a clinically immaterial difference. Pharmacy cost and financial margin differed significantly (both $p < 0.001$): median pharmacy cost was IDR 551,439 (C), 342,097 (A), and 285,397 (B); median margin IDR 309,097 (B), -2,290 (A), and -134,296 (C). After adjustment, median margin was IDR 237,522 and 394,194 lower in A and C than in B, and pharmacy cost 1.18 and 1.71 times higher (all $p < 0.001$). Comparable LOS did not ensure uniform resource consumption; cost variation persisted after adjustment for recorded case-mix and may remain undetected

when efficiency is monitored by hospitalization duration alone. The retrospective design precludes causal attribution. Hospitals should complement LOS monitoring with aggregate physician-group cost audits and digitized clinical pathway governance.

Keywords: clinical pathway; clinical practice variation; financial margin; hospital billing audit; INA-CBG; pharmaceutical cost

PENDAHULUAN

Hospitals operating under universal health coverage increasingly rely on prospective payment systems to control expenditure and promote efficiency. In Indonesia, the INA-CBG payment model reimburses hospitals using predetermined tariffs for diagnosis-based episodes of care. This financing architecture creates a direct operational challenge: when real treatment costs exceed the fixed reimbursement tariff, hospitals absorb the deficit. Accordingly, this study examined whether physician-group variation in pharmaceutical expenditure and total real cost was associated with measurable differences in episode-level financial margin within a single INA-CBG case-mix group, and whether comparable length of stay can be relied upon as an indicator of comparable cost efficiency. Under such conditions, hospital sustainability depends not only on national tariff adequacy but also on internal cost discipline, standardized clinical pathways, and careful monitoring of resource use.

Length of stay (LOS) is frequently used as a proxy indicator for efficiency because it reflects the duration of bed occupancy and service utilization. However, LOS alone may be insufficient to capture the full cost structure of inpatient care. Patients with comparable LOS may still generate very different hospital costs depending on diagnostic intensity, medication use, procedures, consumables, and physician-level clinical decisions. In fixed-tariff systems, these micro-level variations may substantially influence whether an episode produces a surplus or a deficit.

Clinical practice variation is therefore an important but often underexamined determinant of financial performance under prospective payment. While variation may be clinically justified by differences in disease severity, comorbidity, or patient preference, unwarranted variation can lead to inefficient resource utilization. Pharmaceutical expenditure is particularly relevant because it is frequently modifiable through formulary governance, clinical pathway adherence, and prescribing feedback.

This study addresses this gap by examining physician-level variation in pharmaceutical expenditure and financial margin among inpatient episodes within the same INA-CBG category. By focusing on K-4-17-I mild gastroenteritis or abdominal-pain-related episodes and comparing physician groups with broadly comparable LOS, the study evaluates whether similar temporal efficiency is accompanied by similar cost efficiency. Previous evidence shows that reimbursement systems influence resource utilization and care processes, while hospital performance evaluation increasingly requires indicators beyond conventional length-of-stay measures (Hadian et al., 2024; Talebpour et al., 2025; Wagenschieber et al., 2024).

The Interplay of Prospective Payment Systems and Value-Based Healthcare: Prospective payment systems reimburse hospitals using predetermined tariffs for clinically defined groups of care. In Indonesia, the INA-CBG payment model reimburses hospitals based on diagnosis-based episodes of care, creating a direct operational challenge regarding cost containment (Skarayadi et al., 2025; Susilo et al., 2025). When hospitals can deliver appropriate care below the tariff, a positive margin is generated; conversely, when real costs exceed the tariff, hospitals incur a financial loss. Evidence from Medicare's inpatient prospective payment system shows that reimbursement rates themselves vary systematically across providers, so episode-level margin reflects both external tariff setting and internal cost behavior (Krinsky et al., 2017). Indonesian evidence is consistent with this pattern:

when hospital unit costs are computed from actual service activity rather than from historical charges, they frequently diverge from the corresponding INA-CBG package, so the direction and size of the margin differ by procedure, hospital class, and costing method used (Rosmila et al., 2020; Wardani & Pribadi, 2023). In this context, value-based healthcare provides a complementary framework by emphasizing the relationship between clinical outcomes, resource utilization, and cost (Khalil et al., 2025; Porter, 2010; Slingerland et al., 2023). Although prospective payment encourages efficiency, transitioning to value-based healthcare requires robust performance evaluation to ensure that cost reductions do not compromise care quality (He et al., 2024; van Engen et al., 2021). Access has further been argued to be an under-recognized dimension of value, implying that expenditure control must not be pursued at the expense of patient access to appropriate therapy (Chen, 2025).

Deconstructing the Illusion of Temporal Efficiency: Length of stay (LOS) remains a widely used indicator in hospital performance monitoring because it is easy to measure and closely related to bed utilization (Hadian et al., 2024; Talebpour et al., 2025). Nevertheless, LOS alone does not fully explain cost variation. Two patient episodes with the identical duration of stay may differ substantially in medication cost, laboratory use, imaging, and consumables. Therefore, temporal efficiency should be interpreted merely as one component of overall cost efficiency. A more complete evaluation under prospective payment requires an analysis of episode-level micro-costing components, particularly those that are clinically modifiable, such as surgical or pharmaceutical efficiency (Brucchi et al., 2026; Huang et al., 2025; Wang et al., 2025). Activity-based and time-driven activity-based costing studies in Indonesian hospitals reinforce this point, showing that episode cost is determined by the specific bundle of activities and resources consumed rather than by the duration of admission alone (Haq et al., 2023; Nadhiasari et al., 2025).

Physician Practice Variation as the Primary Disruptor of PPS Margins: Physician practice variation refers to systematic differences in clinical decisions that are not fully explained by patient needs or evidence-based guidelines. While some variation is clinically appropriate, unwarranted variation may increase costs without proportional clinical benefit. Prescribing variation is a particularly critical domain because pharmaceutical utilization and prescribing costs can differ considerably across physicians and practices even within comparable clinical contexts (Bucholtz et al., 2018; Kasse et al., 2024; Ntais et al., 2024).

In hospitals where clinical pathways, formulary governance, or prescribing feedback systems are not strictly implemented, physician-level pharmaceutical variation contributes heavily to avoidable cost disparities (Rotter et al., 2010, 2025; Wu et al., 2025). Under a fixed INA-CBG tariff, such micro-level variations can substantially influence whether an episode produces a surplus or a deficit. The novelty of this study lies in its episode-level billing audit of cost divergence within a single INA-CBG code, severity level, clinical department, and broadly comparable LOS profile. By combining unadjusted non-parametric comparisons with covariate-adjusted models, the analysis shows that financial variation remained visible even after recorded case-mix differences were taken into account. This positions pharmaceutical expenditure and total real cost as operational governance indicators that can reveal hidden variation not captured by hospitalization duration alone.

This study aimed to evaluate whether physician-level variation in pharmaceutical expenditure and total real hospital cost is associated with differences in financial margin among inpatient episodes coded as INA-CBG K-4-17-I under a prospective payment system. Specifically, the study compared length of stay, pharmacy cost, total real cost, INA-CBG tariff, and financial margin across three anonymized physician groups to determine whether comparable temporal efficiency is accompanied by comparable cost efficiency.

METODE

This study employed a quantitative retrospective billing-audit design to evaluate physician-level variation in resource utilization and financial outcomes under INA-CBG prospective payment. Data were abstracted from the hospital management information and billing system for inpatient episodes recorded in service year 2025, comprising admissions from 26 December 2024 to 29 December 2025. The target population comprised inpatient episodes categorized under INA-CBG code K-4-17-I, representing mild abdominal pain and gastroenteritis-related cases. This diagnostic group was selected because it is relatively high-volume, clinically common, and suitable for evaluating standardization of care under a fixed prospective tariff.

The study used episode-level administrative and billing data rather than direct clinical trial data. The dataset included case identifier, physician group, attending physician, clinical department, INA-CBG code and description, severity level, admission and discharge dates, LOS, INA-CBG tariff, total real hospital cost, financial margin, age, sex, age group, the coded diagnosis list, the coded procedure list, and the itemised billing components from which pharmacy cost was derived. Ward class, discharge status, and 30-day readmission were not recorded in the billing extract and could therefore not be analysed. Total real hospital cost was defined as the sum of all billed service components recorded against the episode — non-surgical and surgical procedures, consultation, professional and nursing fees, diagnostic support, radiology, laboratory, accommodation, drugs, medical devices, and consumables. Pharmacy cost was defined as the drug expenditure component of that total; medical devices and consumables were counted within total real hospital cost but excluded from the pharmacy variable. Financial margin was calculated per episode as the INA-CBG tariff minus total real hospital cost, and group figures are medians of these episode-level values rather than differences between group medians.

During the study period, 1,600 inpatient episodes were coded as INA-CBG K-4-17-I, all at severity level I, and were managed by 19 attending physicians across ten clinical departments. Of these, 970 episodes (60.6%) were managed by the Department of Internal Medicine. To eliminate specialty mix as a confounder — paediatric and adult gastroenteritis differ systematically in therapeutic regimen and resource use — the analysis was restricted to internal medicine, and within that department to the three highest-volume attending physicians, who together accounted for 88.2% of internal medicine episodes. The final analytical dataset comprised 856 episodes (53.5% of all K-4-17-I episodes): 334 episodes in Group A, 319 in Group B, and 203 in Group C. Each group corresponds to the full caseload of one attending internal medicine physician, so the unit of observation is the inpatient episode while the unit of exposure is the individual prescriber. Physician identities were anonymized to prevent individual attribution, and group-level comparisons were used only for clinical governance and resource utilization analysis.

Inclusion criteria were inpatient episodes coded as INA-CBG K-4-17-I at severity level I during the study period, managed by the Department of Internal Medicine, and attributed to one of the three highest-volume attending physicians in that department. Records were excluded if they were outside the target INA-CBG category or lacked the billing variables required to calculate total real cost, pharmacy cost, tariff, or financial margin. Because this was a retrospective administrative-data audit, no intervention was performed and no patient treatment decisions were influenced by the study.

The primary outcome was financial margin per episode, calculated as INA-CBG tariff minus total real hospital cost. A positive value indicated an operational surplus, whereas a negative value indicated that the hospital incurred a loss for that episode. The main explanatory variable was physician group. Key cost variables included pharmacy cost and total real hospital cost. LOS was assessed to determine whether differences in financial outcome occurred despite comparable hospitalization duration.

Descriptive statistics were used to summarize the number of episodes, LOS, pharmacy cost, total real cost, INA-CBG tariff, and financial margin for each physician group. Continuous variables are reported as median with interquartile range (IQR) given the severe skewness of healthcare cost data, which supports the use of non-parametric analytical frameworks (Conover, 1999; Hollander et al., 2014). The Kruskal-Wallis H test was used to compare LOS, pharmacy cost, total real cost, tariff, and financial margin across the three physician groups. Where the omnibus test was significant, post-hoc pairwise comparisons were performed using Dunn's test with the tie-corrected standard error and Bonferroni adjustment (Dunn, 1964). Tie correction was applied because length of stay takes only six discrete values, so its ranks are extensively tied; omitting the correction inflates the standard error and biases the pairwise tests toward non-significance. Epsilon-squared, calculated as $\epsilon^2 = H / (n - 1)$, was computed for every omnibus test so that the magnitude, and not only the statistical significance, of between-group variation could be judged; values of approximately 0.01, 0.06, and 0.14 were interpreted as small, medium, and large effects, respectively. A two-sided adjusted p-value of < 0.05 was considered statistically significant. All analyses were conducted in Python using pandas and NumPy for data management and descriptive summaries, scipy.stats.kruskal for Kruskal-Wallis testing with correction for tied ranks, and custom mean-rank Dunn pairwise comparisons with Bonferroni adjustment. Because unadjusted rank-based tests cannot exclude confounding by patient characteristics, baseline comparability across the three physician groups was assessed first (age and number of recorded diagnoses by Kruskal-Wallis test; sex and presence of at least one secondary diagnosis by chi-square test), and a covariate-adjusted analysis was then performed as a confirmatory step. Financial margin was modelled by median (quantile) regression at $\tau = 0.50$, which estimates group differences in median margin without assuming normality, and pharmacy cost and total real hospital cost were modelled by generalized linear models with a gamma distribution and log link, the conventional specification for right-skewed cost data. All three models included physician group as the exposure with Group B as the reference category and were adjusted for age, sex, presence of at least one secondary diagnosis, and length of stay; adjusted results are reported as differences in median margin with 95% confidence intervals and as cost ratios with 95% confidence intervals. Case-mix confounding was additionally controlled by design, since the cohort was restricted to a single INA-CBG code (K-4-17-I), a single severity level (I), and a single clinical department, so that specialty and tariff-band heterogeneity were held constant by construction. Regression models were fitted in Python with statsmodels. The analysis nevertheless emphasises association rather than causality because the study is retrospective and observational, and residual confounding from unmeasured within-stratum variation, including true disease presentation and clinical coding practice, cannot be excluded.

This study was reviewed and approved by the Health Research Ethics Committee of the Faculty of Medicine and Health Sciences, Universitas Muhammadiyah Yogyakarta, with ethical clearance number No.094/EC-EXEM-KEPK FKIK UMY/VII/2026. The study used retrospective administrative and billing data that had been anonymized before analysis. Patient identifiers were not reported, and physician identities were coded as Physician A, Physician B, and Physician C in the manuscript to prevent individual attribution. The analysis was intended as a clinical governance and resource utilization audit rather than a personal performance assessment of individual physicians. Results are presented in aggregate form to support quality improvement, clinical pathway monitoring, and financial sustainability under prospective payment.

HASIL

To maintain ethical standards and data confidentiality, the identities of all attending physicians were strictly anonymized into three cohorts: Group A, Group B, and Group C. The final retrospective analysis included 856 inpatient episodes categorized under INA-CBG K-4-17-I. The distribution of cases was as follows: Group A contributed 334 episodes, Group B contributed 319 episodes, and Group C contributed 203 episodes. The cohort comprised adult patients with a median age of 48 years (range 17–94) and was 61.1% female (523 of 856); all episodes were coded at severity level I.

Length of Stay Comparison: The median length of stay (LOS) was identical across all three physician groups at 3 days, with an interquartile range (IQR) of 3–4 days (see Table 1). The omnibus Kruskal-Wallis test detected a statistically significant global difference in LOS ($H = 8.49$, $p = 0.014$), but the associated effect size was negligible ($\epsilon^2 = 0.010$), and the median duration was 3 days in every group. After tie-corrected Dunn testing with Bonferroni adjustment, Group C exhibited significantly longer stays than Group B ($Z = -2.883$, adjusted $p = 0.012$), whereas neither comparison involving Group A reached significance (A vs B, adjusted $p = 1.000$; A vs C, adjusted $p = 0.105$). Taken together with the negligible effect size, the statistically detectable difference in LOS between Group B and Group C is of negligible magnitude: all three groups shared a median stay of 3 days with an identical interquartile range, and the LOS effect size is approximately one-fourteenth of that observed for pharmacy cost and financial margin. Temporal variation was therefore present but immaterial relative to cost variation.

Table 1.
Median, Interquartile Range, and Kruskal-Wallis Results by Physician Group

Variable	Group A, median (IQR)	Group B, median (IQR)	Group C, median (IQR)	Kruskal-Wallis p -value	H (ϵ^2)
LOS, days	3 (3–4)	3 (3–4)	3 (3–4)	0.014	8.49 (0.010)
Pharmacy cost, IDR	342,097 (224,292–515,667)	285,397 (196,047–430,355)	551,439 (372,397–786,565)	<0.001	120.68 (0.141)
Total real cost, IDR	1,687,270 (1,387,654–2,145,553)	1,421,856 (1,200,508–1,842,228)	1,835,786 (1,541,517–2,418,856)	<0.001	77.12 (0.090)
INA-CBG tariff, IDR	1,635,100 (1,635,100–1,904,900)	1,635,100 (1,635,100–1,904,900)	1,635,100 (1,635,100–1,904,900)	0.052	5.90 (0.007)
Financial margin, IDR	-2,290 (-343,098–292,990)	309,097 (42,769–493,740)	-134,296 (-586,983–159,689)	<0.001	121.35 (0.142)

Cost Variation: Despite the uniform temporal efficiency, micro-costing components revealed substantial between-group differences. Pharmacy costs differed significantly across the physician groups, with a large effect size ($H = 120.68$, $p < 0.001$, $\epsilon^2 = 0.141$) — approximately fourteen times the magnitude of the LOS effect. The median pharmacy cost was highest in Group C (IDR 551,439), followed by Group A (IDR 342,097), and was lowest in Group B (IDR 285,397). Total real hospital costs mirrored this trajectory, with Group C generating the highest median total real cost (IDR 1,835,786) and Group B the lowest (IDR 1,421,856). In stark contrast, the INA-CBG reimbursement tariffs did not differ significantly across groups ($p = 0.052$), indicating that external reimbursement rates were statistically comparable across the groups. The tariff distribution was in fact identical across the three groups at every reported quantile (median IDR 1,635,100; Q1 IDR 1,635,100; Q3 IDR 1,904,900), reflecting the small number of discrete tariff values available within a single severity level. Because the reported tariff quantiles are identical across groups, the borderline p -value should not be interpreted as evidence of a practically meaningful difference in reimbursement. Reimbursement can therefore be treated as effectively fixed across groups, which isolates internal cost behavior as the source of margin variation.

Margin Outcome: Consequently, the financial margin differed significantly across the groups ($p < 0.001$). Group B achieved the highest median positive margin (IDR 309,097), generating a surplus for the hospital. Group A operated near break-even with a median margin of IDR -2,290, whereas Group C recorded the largest median financial deficit, presenting a median margin of IDR -134,296. The between-group difference in financial margin was large ($H = 121.35$, $p < 0.001$, $\epsilon^2 = 0.142$), closely tracking the magnitude of the pharmacy cost effect. This correspondence indicates that variation in real cost, predominantly pharmaceutical, accounts for most of the observed divergence

in financial outcome under identical INA-CBG limits.

Baseline Comparability: The three groups were not fully comparable at baseline (see Table 2). Median age was 50 years (IQR 36–60) in Group A, 43 years (IQR 27–56) in Group B, and 49 years (IQR 34–59) in Group C ($p < 0.001$), and the proportion of episodes with at least one secondary diagnosis was 32.6% (109/334) in Group A, 12.5% (40/319) in Group B, and 32.0% (65/203) in Group C ($p < 0.001$). The proportion of female patients did not differ significantly (56.3%, 63.9%, and 64.5% respectively; $p = 0.069$), and median LOS was 3 days (IQR 3–4) in every group. Because Group B — the group with the most favourable financial margin — also had the youngest patients and the lowest comorbidity burden, covariate adjustment was required before the observed cost and margin differences could be attributed to practice variation.

Table 2.

Baseline Characteristics of Inpatient Episodes by Physician Group

Variable	Group A (n = 334)	Group B (n = 319)	Group C (n = 203)	p- value
Age, years, median (IQR)	50 (36–60)	43 (27–56)	49 (34–59)	<0.001
Female sex, n (%)	188 (56.3)	204 (63.9)	131 (64.5)	0.069
Episodes with ≥ 1 secondary diagnosis, n (%)	109 (32.6)	40 (12.5)	65 (32.0)	<0.001
Number of coded diagnoses, median (IQR)	1 (1–2)	1 (1–1)	1 (1–2)	<0.001
Length of stay, days, median (IQR)	3 (3–4)	3 (3–4)	3 (3–4)	0.014

Note. Age and number of coded diagnoses compared by Kruskal-Wallis test; sex and presence of ≥ 1 secondary diagnosis by chi-square test.

Adjusted Analysis: The between-group differences persisted after adjustment for age, sex, presence of at least one secondary diagnosis, and length of stay (see Table 3). In median regression with Group B as the reference, the adjusted median financial margin was IDR 237,522 lower in Group A (95% CI 162,186 to 312,857 lower; $p < 0.001$) and IDR 394,194 lower in Group C (95% CI 308,670 to 479,719 lower; $p < 0.001$). In gamma-log generalized linear models, adjusted pharmacy cost was 1.18 times higher in Group A (95% CI 1.08–1.29; $p < 0.001$) and 1.71 times higher in Group C (95% CI 1.55–1.89; $p < 0.001$) than in Group B, and adjusted total real hospital cost was 1.13 times higher in Group A (95% CI 1.07–1.20; $p < 0.001$) and 1.19 times higher in Group C (95% CI 1.12–1.27; $p < 0.001$). Among the covariates, each additional inpatient day was associated with an IDR 412,039 reduction in median margin ($p < 0.001$) and the presence of a secondary diagnosis with a further IDR 131,253 reduction ($p < 0.001$), whereas age and sex were not significantly associated with margin. Physician group therefore remained an independent correlate of pharmacy cost, total real cost, and financial margin after adjustment for the recorded case-mix differences.

Table 3.

Covariate-Adjusted Comparison of Cost and Margin Outcomes by Physician Group

Outcome (model)	Effect measure	Group A vs Group B (95% CI)	Group C vs Group B (95% CI)	p-value
Financial margin (median regression)	Difference in median margin, IDR	-237,522 (-312,857 to - 162,186)	-394,194 (-479,719 to - 308,670)	<0.001
Pharmacy cost (gamma GLM, log link)	Cost ratio	1.18 (1.08–1.29)	1.71 (1.55–1.89)	<0.001
Total real hospital cost (gamma GLM, log link)	Cost ratio	1.13 (1.07–1.20)	1.19 (1.12–1.27)	<0.001

Note. Group B is the reference category. All models are adjusted for age, sex, presence of at least one secondary diagnosis, and length of stay. Within the margin model, each additional inpatient day was associated with an IDR 412,039 lower median margin ($p < 0.001$) and the presence of at least one secondary diagnosis with an IDR 131,253 lower median margin ($p < 0.001$); age and sex were not significantly associated with margin.

Post-Hoc Pairwise Analysis: To evaluate specific group differences, Dunn's post-hoc test with Bonferroni correction was conducted (see Table 4).

For Length of Stay (LOS), only the Group B versus Group C comparison remained statistically significant after the correction (adjusted p-values: A vs B = 1.000; A vs C = 0.105; B vs C = 0.012), which, combined with the negligible omnibus effect size, indicates that no clinically meaningful difference in hospitalization duration was detected between any pair of groups.

Table 4.

Dunn Post-hoc Pairwise Comparisons with Bonferroni Correction				
Outcome	Comparison	Z-statistic	Adjusted p-value	Interpretation
LOS	Group A vs Group B	0.910	1.000	Not significant
LOS	Group A vs Group C	-2.109	0.105	Not significant
LOS	Group B vs Group C	-2.883	0.012	Group C longer LOS than Group B
Pharmacy cost	Group A vs Group B	3.616	0.001	Group A higher than Group B
Pharmacy cost	Group A vs Group C	-7.813	<0.001	Group C higher than Group A
Pharmacy cost	Group B vs Group C	-10.897	<0.001	Group C higher than Group B
Total real cost	Group A vs Group B	6.328	<0.001	Group A higher than Group B
Total real cost	Group A vs Group C	-2.779	0.016	Group C higher than Group A
Total real cost	Group B vs Group C	-8.272	<0.001	Group C higher than Group B
Financial margin	Group A vs Group B	-8.079	<0.001	Group B higher than Group A
Financial margin	Group A vs Group C	3.291	0.003	Group A higher than Group C
Financial margin	Group B vs Group C	10.306	<0.001	Group B higher than Group C

Conversely, all pairwise comparisons for pharmacy costs and total real costs were statistically significant (with adjusted p-values ranging from $p < 0.001$ to $p = 0.016$), establishing a consistent cost hierarchy of Group C > Group A > Group B. Pharmacy cost was highest in Group C, intermediate in Group A, and lowest in Group B. Regarding the financial margin, all post-hoc pairwise comparisons were also statistically significant, but the direction was reversed. Because the groups operated under identical INA-CBG reimbursement packages, the higher pharmacy expenditure of Groups A and C is mirrored one-for-one in their lower margins. Group B consistently and significantly outperformed Group A, while Group C yielded the lowest median margin.

PEMBAHASAN

This retrospective billing-data-driven audit indicates that comparable length of stay (LOS) does not necessarily imply comparable cost efficiency under INA-CBG prospective payment. While median LOS was uniformly 3 days across all anonymized physician groups, median financial margin ranged from a surplus of IDR 309,097 in Group B to a deficit of IDR 134,296 in Group C. Because the INA-CBG tariff was identical in median terms and did not differ significantly across the groups, the divergence in margin is located in internal resource consumption rather than in reimbursement inequality; why resource use differed cannot be established from billing data alone. The pairwise comparisons in Table 4 indicate that micro-level resource use, and pharmaceutical expenditure in particular, is closely associated with financial performance in fixed-tariff settings. Group C recorded both the highest median pharmacy cost (IDR 551,439) and the lowest median financial margin.

Group B recorded the lowest median pharmacy expenditure (IDR 285,397) and the only positive median margin. This inverse relationship suggests that even when temporal efficiency (LOS) is identical, variation in prescribing patterns may accompany materially different financial outcomes at the episode level. This pattern is consistent with health economics literature indicating that standardized prescribing, formulary governance, and rigorous clinical pathways are essential to control hospital expenditure effectively (Buchholz et al., 2018; Kasse et al., 2024; Rotter et al., 2025; Wu et al., 2025). Pharmaceutical spending is additionally shaped by acquisition and negotiated price structures that differ markedly between institutions, which reinforces the case for treating drug expenditure as a governable rather than a fixed cost component (Feldman et al., 2022; Robinson et al., 2024).

An alternative explanation for this pattern is case-mix: Group B treated the youngest patients and had the lowest proportion of episodes with a secondary diagnosis, so its favourable margin could in principle reflect an easier caseload rather than more economical practice. The covariate-adjusted models argue against that reading. After adjustment for age, sex, secondary diagnosis, and length of stay, pharmacy cost remained 1.18 times higher in Group A and 1.71 times higher in Group C than in Group B, and adjusted median margin remained IDR 237,522 and IDR 394,194 lower respectively, with confidence intervals excluding the null in every model. Case-mix does exert a measurable effect — each additional inpatient day and the presence of a secondary diagnosis were each independently associated with a lower margin — but it accounts for only part of the gap. The residual, physician-level component is therefore the plausible target for clinical governance rather than for tariff renegotiation.

In the context of Indonesia's National Health Insurance (JKN) system, these findings point to a gap in standard hospital performance metrics. Hospital quality and cost-control committees frequently rely on macro-indicators like LOS to proxy clinical efficiency. However, this study shows that material cost differences can be present without appearing in length-of-stay indicators, a systemic challenge documented in evaluations of INA-CBG tariff adequacy and real-cost discrepancies (Sari et al., 2024; Skarayadi et al., 2025; Susilo et al., 2025; Syakurah et al., 2025). The same vulnerability appears at the institutional level under payment-reform pilots, where hospitals that had not conducted tariff, medical-service-fee, or medication-cost reviews recorded recurrent gaps between projected budgets and realized claims (Alvianto & Pribadi, 2025), suggesting that financial exposure is associated with the absence of internal cost review and not with the payment mechanism alone. Variation in pharmaceutical expenditure—which may reflect differing adherence to the National Formulary or to local clinical pathways—was the cost component most strongly associated with margin divergence in this cohort. Experience from other low- and middle-income settings indicates that adapting global prescribing guidance to local formularies, supported by audit-and-feedback mechanisms, can measurably reduce unwarranted prescribing variation (Jamil et al., 2025). Under prospective payment, the financial consequences of prescribing variation are borne by the institution rather than the payer, which places clinical governance at the center of cost control (Ntais et al., 2024; Rotter et al., 2025).

Managerial Implications: The results have several practical implications for hospital managers implementing value-based healthcare (He et al., 2024; Khalil et al., 2025). First, hospital administrators must transition away from relying solely on LOS as a proxy for efficiency; high-volume INA-CBG groups must be evaluated using granular cost-component indicators. Second, the integration of lean and six sigma strategies within pharmacy operations can help standardize medication use and reduce unwarranted expenditures (Sallam, 2024). Third, the cost-efficient practices demonstrated by Group B can serve as an internal benchmark for peer-to-peer, non-punitive clinical governance reviews. Finally, implementing computerized provider order entry systems with formulary prompts is recommended to correct non-standard prescribing habits dynamically at the point of care.

Limitations: This study has several limitations. First, the retrospective single-center design inherently limits the generalizability of the findings to broader hospital settings. Second, although the analysis was restricted to a single INA-CBG category and severity level and was additionally adjusted for age, sex, secondary diagnosis, and length of stay, residual confounding from unmeasured factors—including true disease presentation, ward class, and clinical coding practice, none of which were recorded in the billing extract—may remain. Third, the reliance on administrative billing data may obscure nuanced clinical contexts, including true disease severity, patient preferences, and medication availability at the point of care. Fourth, while the analysis robustly demonstrates associations between physician group, pharmaceutical expenditure, and financial margin, it is observational and cannot establish definitive causality. Fifth, no clinical outcome or quality-of-care measures — such as treatment success, complication rate, mortality, or 30-day readmission — were analyzed, so lower pharmaceutical expenditure cannot be interpreted as superior clinical performance; the possibility that the lowest-cost prescribing pattern reflects under-treatment rather than efficiency remains untested. Sixth, each comparison group represents the caseload of a single attending physician, so episodes are clustered within three prescribers. The analysis treated episodes as independent observations and did not model this clustering; consequently the p-values should be read as describing differences between these three specific practice patterns rather than as generalizable estimates of prescriber-level variation, and the effective number of independent exposure units is three rather than 856. Seventh, restricting the cohort to severity level I and to internal medicine strengthens internal comparability but limits extrapolation to pediatric cases, higher severity levels, and other specialties managing the same INA-CBG code. Future studies should incorporate patient-level clinical variables beyond those available in billing records, mixed-effects models that account for clustering within prescribers, outlier sensitivity analysis, paired cost–quality outcome measures, and multicenter datasets to further validate these findings.

SIMPULAN

This retrospective billing audit confirms that within the same diagnostic category (INA-CBG K-4-17-I), statistically significant variations exist in pharmacy costs, total real hospital costs, and resulting financial margins across different physician groups. Although the median length of stay was identical at 3 days in all three groups and the LOS effect size was negligible ($\varepsilon^2 = 0.010$), financial outcomes diverged substantially. Pharmacy and total real costs differed significantly across all pairwise comparisons, being highest in Group C and lowest in Group B. This directly impacted the financial margins, where Group B achieved the highest median surplus and Group C incurred the largest median deficit. These findings indicate that micro-costing variation in pharmaceutical expenditure is the dominant correlate of financial margin under prospective payment, and that uniform temporal indicators such as LOS may mask substantial cost heterogeneity within a single case-mix group. Because the financial consequences of individual prescribing decisions are borne by the hospital rather than the payer, reliance on macro-level performance indicators alone leaves this source of variation unmonitored. To safeguard financial viability under JKN prospective limits, hospital management must implement proactive, aggregate physician-level billing audits, strengthen the enforcement of local clinical pathways, and integrate real-time formulary governance protocols.

DAFTAR PUSTAKA

- Alvianto, M. J., & Pribadi, F. (2025). Analysis of Indonesian Case-Based Groups with global budget pilot project implementation. *Indonesian Journal of Innovation Studies*, 26(4). <https://doi.org/10.21070/ijins.v26i4.1419>
- Brucchi, F., Colombo, C., Bianchi, P. P., Mercante, G., Barbieri, D., Lombardi, D., Spriano, G., Bussi, M., Giordano, L., Boni, L., Danelli, P., Persani, L., Materazzi, G., Miccoli, P., Dionigi, R., & Dionigi, G. (2026). ERAS thyroidectomy within the Italian DRG-based public health

- system: Surgical efficiency is penalized. *Updates in Surgery*. <https://doi.org/10.1007/s13304-026-02698-4>
- Bucholtz, M., O’Kane, M., Ashe, S., & Wong-Lin, K. (2018). Prescriptive variability of drugs by general practitioners. *PLoS ONE*, 13(2), e0189599. <https://doi.org/10.1371/journal.pone.0189599>
- Chen, A. M. (2025). Why access matters in value-based healthcare: A systematic review. *Journal for Healthcare Quality*. <https://doi.org/10.1097/JHQ.0000000000000471>
- Conover, W. J. (1999). *Practical Nonparametric Statistics* (3rd ed.). Wiley.
- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6(3), 241–252. <https://doi.org/10.1080/00401706.1964.10490181>
- Feldman, W. B., Rome, B. N., Brown, B. L., & Kesselheim, A. S. (2022). Payer-specific negotiated prices for prescription drugs at top-performing US hospitals. *JAMA Internal Medicine*, 182(1), 83–86. <https://doi.org/10.1001/jamainternmed.2021.6445>
- Hadian, S. A., Rezayatmand, R., Shaarbafchizadeh, N., Ketabi, S., & Pourghaderi, A. R. (2024). Hospital performance evaluation indicators: A scoping review. *BMC Health Services Research*, 24, 566. <https://doi.org/10.1186/s12913-024-10940-1>
- Haq, F. A., Pribadi, F., & Akrom, A. (2023). Elective cesarean section cost efficiency with time-driven activity-based costing. *International Journal of Public Health Science*, 12(1), 155–163.
- He, W., Feng, W., Du, L., & Zhang, W. (2024). Key factors in the implementation of value-based healthcare performance evaluation in general hospital healthcare groups: A mixed methodology study. *Journal of Clinical Nursing*. <https://doi.org/10.1111/jocn.17597>
- Hollander, M., Wolfe, D. A., & Chicken, E. (2014). *Nonparametric Statistical Methods* (3rd ed.). Wiley.
- Huang, X., Zhu, J., & Zhang, J. (2025). Impacts of DRG point-based payment system on healthcare resource utilization and provider behavior: A pilot quasi-experimental study in China. *Frontiers in Public Health*, 13, 1678259. <https://doi.org/10.3389/fpubh.2025.1678259>
- Jamil, E., Aziz, M. M., Amir, A., Godman, B., Campbell, S. M., Ullah, M., Arooj, H., Altowayan, W., & Saleem, Z. (2025). Adapting global guidelines to local contexts: Optimising community-acquired pneumonia-specific prescribing in Pakistan to counter antimicrobial resistance. *BMJ Open Respiratory Research*. <https://doi.org/10.1136/bmjresp-2025-003699>
- Kasse, G., Humphries, J., Cosh, S. M., & Islam, M. S. (2024). Factors contributing to the variation in antibiotic prescribing among primary health care physicians: A systematic review. *BMC Primary Care*, 25, 1. <https://doi.org/10.1186/s12875-023-02223-1>
- Khalil, H., Ameen, M., Davies, C. C., & Liu, C. (2025). Implementing value-based healthcare: A scoping review of key elements, outcomes, and challenges for sustainable healthcare systems. *Frontiers in Public Health*, 13, 1514098. <https://doi.org/10.3389/fpubh.2025.1514098>
- Krinsky, S., Ryan, A. M., Mijanovich, T., & Blustein, J. (2017). Variation in payment rates under Medicare’s inpatient prospective payment system. *Health Services Research*, 52(2), 676–696. <https://doi.org/10.1111/1475-6773.12490>
- Nadhiasari, A., Pribadi, F., & Nazaruddin, I. (2025). Unit cost analysis in heart failure inpatients using the time-driven activity-based costing method at a type D hospital in Sragen. *Health Policy and Management*, 10(1), 1–15.
- Ntais, C., Talias, M. A., Fanourgiakis, J., & Kontodimopoulos, N. (2024). Managing pharmaceutical costs in health systems: A review of affordability, accessibility and sustainability strategies. *Journal of Market Access & Health Policy*, 12(4), 403–414. <https://doi.org/10.3390/jmahp12040031>
- Porter, M. E. (2010). What is value in health care? *New England Journal of Medicine*, 363(26), 2477–2481. <https://doi.org/10.1056/NEJMp1011024>
- Robinson, J. C., Whaley, C., & Dhruva, S. S. (2024). Hospital prices for physician-administered drugs for patients with private insurance. *New England Journal of Medicine*, 390, 338–345. <https://doi.org/10.1056/NEJMs2306609>
- Rosmila, G. B., Yaya, R., & Pribadi, F. (2020). The comparison of the unit cost of hemodialysis

- with INA-CBG rates in Muhammadiyah Siti Aminah Hospital. *Archives of Business Research*, 8(5), 80–95. <https://doi.org/10.14738/abr.85.8210>
- Rotter, T., Kinsman, L., James, E., Machotta, A., Gothe, H., Willis, J., Snow, P., & Kugler, J. (2010). Clinical pathways: Effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database of Systematic Reviews*, CD006632. <https://doi.org/10.1002/14651858.CD006632.pub2>
- Rotter, T., Kinsman, L. D., Alsius, A., Scott, S. D., Lawal, A., Ronellenfitsch, U., Plishka, C., Groot, G., Woods, P., Coulson, C., Bakel, L. A., Sears, K., Ross-White, A., Machotta, A., & Schultz, T. J. (2025). Clinical pathways for secondary care and the effects on professional practice, patient outcomes, length of stay and hospital costs. *Cochrane Database of Systematic Reviews*, 5, CD006632. <https://doi.org/10.1002/14651858.CD006632.pub3>
- Sallam, M. (2024). Enhancing hospital pharmacy operations through lean and six sigma strategies: A systematic review. *Cureus*, 16(3), e57176. <https://doi.org/10.7759/cureus.57176>
- Sari, N. M., Sulaiman, M. H., Khairunisaa, Azzahra, N. A., Napirah, M. R., & Wardana Toii, I. E. (2024). Tariff of Indonesia Case Based Groups (INA-CBG's) and real tariff on acute gastroenteritis patient national health insurance in inpatient unit class I district general hospital Undata Palu. *International Journal of Health, Economics, and Social Sciences*, 6(4), 1173–1178. <https://doi.org/10.56338/ijhess.v6i4.6003>
- Skarayadi, O., Endarti, D., Satibi, S., & Mukti, A. G. (2025). Discrepancies in hospital cost and Indonesia's Case-based Groups: A literature review. *Jurnal Manajemen dan Pelayanan Farmasi*. <https://doi.org/10.22146/jmpf.95970>
- Slingerland, S., Schulz, D., Dekker, L., van der Wees, P., van der Meer, N. J., Groot, W., & van Veghel, D. (2023). The role of value-based healthcare in decision-making by executives in hospitals in the Netherlands: A nationwide evaluation. *International Journal of Healthcare Management*. <https://doi.org/10.1080/20479700.2023.2252643>
- Susilo, D., Wulandari, L., Sukmayeti, E., Asante, A., Jan, S., Thabrany, H., Tangcharoensathien, V., Wiseman, V., & Liverani, M. (2025). Can Indonesia achieve universal health coverage? Organisational and financing challenges in implementing the national health insurance system. *SSM - Health Systems*. <https://doi.org/10.1016/j.ssmhs.2025.100138>
- Syakurah, R. A., Frymaruwah, E., Mufarikah, M., & Izzatika, M. (2025). Comparative analysis of INA-CBGs tariffs and unit cost of cesarean section: A systematic review in Indonesian hospitals. *Journal of Public Health Sciences*, 4(3). <https://doi.org/10.56741/IISTR.jphs.001257>
- Talebpour, A., Sadeghi-Bazargani, H., Janati, A., Pashazadeh, F., & Gholizadeh, M. (2025). Crucial key performance indicators for hospital evaluation: A scoping review. *Journal of Education and Health Promotion*, 14, 195. https://doi.org/10.4103/jehp.jehp_2102_23
- van Engen, V., Bonfrer, I., Ahaus, K., & Buljac-Samardzic, M. (2021). Value-based healthcare from the perspective of the healthcare professional: A systematic literature review. *Frontiers in Public Health*, 9, 800702. <https://doi.org/10.3389/fpubh.2021.800702>
- Wagenschieber, E., Blunck, D., & Busse, R. (2024). Impact of reimbursement systems on patient care: A systematic review of systematic reviews. *Health Economics Review*, 14, 22. <https://doi.org/10.1186/s13561-024-00487-6>
- Wang, J., Zhu, J., Hu, K., Chen, Y., Zhang, J., & Wu, X. (2025). Cost shifting or cost cutting by hospitals as a response to reimbursement reform? The case of diagnosis-related groups scheme in China. *Frontiers in Public Health*, 13, 1582001. <https://doi.org/10.3389/fpubh.2025.1582001>
- Wardani, D. T. K., & Pribadi, F. (2023). Unit cost calculation of elective caesarean section without complications based on the time-driven activity-based costing model. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 8(2), 9–16. <https://doi.org/10.30604/jika.v8i2.1824>
- Wu, S. Q., Wang, X. C., Boyd, A. D., Feng, D., Zhong, M., & Nie, D. (2025). Exploration of clinical pathway practice for optimization of DRG costing results based on resource consumption. *BMC Medical Informatics and Decision Making*, 25, 305. <https://doi.org/10.1186/s12911-025-03152-y>