

“Kazakhstan healthcare reforms in the context of public administration: Analysis of results and public opinion”

AUTHORS

Talgat Zharlygassinov 
Dina Razakova 

Aigerim Sekerbayeva 


ARTICLE INFO

Talgat Zharlygassinov, Dina Razakova and Aigerim Sekerbayeva (2026). Kazakhstan healthcare reforms in the context of public administration: Analysis of results and public opinion. *Problems and Perspectives in Management*, 24(4), 69–85. doi:[10.21511/ppm.24\(4\).2026.05](https://doi.org/10.21511/ppm.24(4).2026.05)

DOI [http://dx.doi.org/10.21511/ppm.24\(4\).2026.05](http://dx.doi.org/10.21511/ppm.24(4).2026.05)

RELEASED ON Tuesday, 13 October 2026

RECEIVED ON Wednesday, 03 December 2025

ACCEPTED ON Tuesday, 25 August 2026

LICENSE



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

JOURNAL

"Problems and Perspectives in Management"

ISSN PRINT

1727-7051

ISSN ONLINE

1810-5467

PUBLISHER

LLC "Consulting Publishing Company "Business Perspectives"

FOUNDER

LLC "Consulting Publishing Company "Business Perspectives"



NUMBER OF REFERENCES

36



NUMBER OF FIGURES

3



NUMBER OF TABLES

4

© The author(s) 2026. This publication is an open access article.



BUSINESS PERSPECTIVES



LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine
www.businessperspectives.org

Type of the article: Research Article

Received on: 3rd of December, 2025

Accepted on: 25th of August, 2026

Published on: 13th of October, 2026

© Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva, 2026

Talgat Zharlygassinov, Postdoctoral Student, Higher School of Economics and Business, Al-Farabi Kazakh National University, Republic of Kazakhstan.

Dina Razakova, Ph.D., Professor, Vice-Rector for Strategic Development, Science and Innovation, Department of Marketing and Logistics, Faculty of Economics, Turan University, Republic of Kazakhstan.

Aigerim Sekerbayeva, Ph.D., Assistant Professor, School of Economics and Management, Narxoz University, Republic of Kazakhstan. (Corresponding author)



This is an Open Access article, distributed under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license, which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

Conflict of interest statement:

Author(s) reported no conflict of interest

Talgat Zharlygassinov (Republic of Kazakhstan), Dina Razakova (Republic of Kazakhstan), Aigerim Sekerbayeva (Republic of Kazakhstan)

KAZAKHSTAN HEALTHCARE REFORMS IN THE CONTEXT OF PUBLIC ADMINISTRATION: ANALYSIS OF RESULTS AND PUBLIC OPINION

Abstract

Healthcare reforms in transitional economies frequently prioritize administrative performance indicators; however, citizen evaluations often diverge significantly from these official metrics. To explore this discrepancy, this study investigates the extent to which reported healthcare improvements in Kazakhstan align with actual patient experiences, examining how geographic and socioeconomic factors shape the performance-perception gap. Utilizing a mixed-methods approach, the paper integrates documentary analysis of policy materials (2019–2024), regional health statistics (2019–2023), and a nationwide stratified survey ($n = 1,940$) conducted across all 20 regions of Kazakhstan in 2024.

Results demonstrate a persistent, statistically significant divergence between official gains, such as increased life expectancy and infrastructure modernization, and public sentiment. Specifically, only 28.35% of respondents report high service quality (7.73% very high + 20.62% high), and just 28.7% express satisfaction levels above the neutral midpoint. Regression and ANOVA analyses confirm that disparities in access and affordability are systematically patterned by age, income, and urban-rural residence ($p < 0.05$). This gap is most pronounced in resource-constrained regions, where administrative progress in financing and digitalization has not translated into improved patient experiences. Further analysis identifies waiting times, transparency deficits, and perceived inequity as the primary drivers of this divergence. Based on these findings, the study concludes that existing evaluative methods are inadequate, suggesting that integrating standardized administrative metrics with regionally disaggregated citizen feedback is essential to align governance interventions with public expectations for the 2026 healthcare development horizon.

Keywords

healthcare reforms, public administration, patient experience, accessibility, affordability, performance measurement, equity, Kazakhstan

JEL Classification

I15, I18, H75, H83

INTRODUCTION

Evaluating healthcare reform in transitional economies requires a critical perspective that moves beyond traditional administrative metrics. In post-Soviet contexts, a significant performance paradox often emerges: improvements in official indicators do not translate into better healthcare experiences for citizens. While official indicators signal technical progress in financing and infrastructure, citizens still face persistent barriers to access and quality. This divergence suggests a state of institutional decoupling, meaning formal administrative successes are disconnected from actual care delivery. Understanding this gap is not merely a statistical exercise but a fundamental governance concern.

Kazakhstan is a compelling case for this study. Since 2019, the country has implemented several major reform programs. These include

the State Health Development Programme 2020–2025 (Government Resolution No. 982; Government of the Republic of Kazakhstan, 2019), which targeted financing reform, primary care strengthening, and reduction of out-of-pocket payments. Next, the Concept for Healthcare Development until 2026 (Government Resolution No. 945, amended by Resolution No. 141; Government of the Republic of Kazakhstan, 2024) set system-wide quality and equity targets. The mandatory social health insurance framework (OSMS) was fully operationalized in 2020, along with a suite of digital health initiatives including e-Denisaul (electronic health records), the Damumed telemedicine platform, and the integrated electronic polyclinic system (IEPC). Despite official reports of system growth across these programs, public concerns about transparency and affordability remain. Current evaluation frameworks in Kazakhstan remain fragmented. They rarely integrate macro-level administrative data with micro-level patient experiences within one subnational framework. As a result, the spatial and social dimensions of the performance–perception gap remain poorly understood.

1. LITERATURE REVIEW AND HYPOTHESES

The divergence between administrative performance assessments and citizen-reported healthcare experience constitutes an established theoretical and empirical concern across public administration, health policy, and organizational research. This review synthesizes evidence across three inter-related dimensions directly relevant to the present study: governance mechanisms mediating reform translation into service delivery; distributional dynamics of accessibility and affordability under reform conditions; and patient satisfaction and perceived quality as governance-relevant outcomes.

1.1. Governance, implementation capacity, and the reform–experience gap

Healthcare reform outcomes in transitional systems are not determined by policy design alone. Public management scholarship emphasizes that reforms produce service-level changes through implementation capacity, frontline governance, and feedback mechanisms rather than through formal legislative change in isolation (Ferlie & McGivern, 2014; Newman & Lawler, 2009). The diffusion of managerial instruments across institutional settings is inherently uneven, generating heterogeneity in outcomes even when reform packages are nationally standardized (Vamstad & Gordon, 2018). This perspective implies that administrative performance indicators (however technically valid) are imperfect proxies for lived service experience, and that perceived quality and satisfaction constitute governance-relevant outcomes.

A well-documented phenomenon in this literature is the performance paradox: an observed weakening of the link between reported indicators and underlying organizational performance as measurement systems become embedded in administrative practice (van Thiel & Leeuw, 2002). Under these conditions, official metrics may overstate improvements in actual service delivery, and the correspondence between administrative data and citizen-reported experience cannot be assumed but must be empirically tested (Van de Walle, 2018). Comparative studies across post-Soviet health systems suggest that population satisfaction can diverge markedly from official performance signals, reinforcing the need for integrated evaluation frameworks rather than single-metric assessments (Footman et al., 2013).

A specific organizational mechanism through which this divergence materializes is institutional decoupling: the adoption of formal structures, reporting standards, or compliance procedures that signal improvement without generating commensurate changes in frontline practice (Meyer & Rowan, 1977; Bromley & Powell, 2012). In healthcare, this dynamic can produce measurable gains in administrative indicators (infrastructure investment, financing inputs, or mortality statistics), while waiting times, communication quality, and perceived responsiveness remain unchanged or deteriorate. This mechanism is particularly salient in centralized governance settings, where upward accountability to administrative targets may crowd out horizontal responsiveness to users.

Digital transformation, increasingly central to reform strategies in transitional systems, reproduces

these same governance dynamics. Evidence indicates that eHealth tools, electronic registries, and digital service pathways improve access and user experience only when usability, workflow integration, and organizational adoption are achieved (Andersen, 2019; Vamstad & Gordon, 2018). Conversely, digitalization can generate administrative modernization measurable in investment and platform deployment without proportionate benefit for users if system capacity, staff training, or service design is insufficient. Critically, digital pathways also introduce equity risks: digital access and digital literacy condition who can benefit from digitized service routes, shaping unequal ability to register, navigate entitlements, or access reliable health information (Alhrahsheh & Ivanova, 2022; Suárez-Brito et al., 2024). In contexts characterized by pronounced urban-rural and income gradients, digital channels may therefore widen disparities even as aggregate system indicators improve.

1.2. Accessibility, affordability, and distributional dynamics of reform

International comparative evidence consistently demonstrates that coverage expansion, infrastructure investment, and financing reforms do not automatically translate into equitable access or reduced financial burden. Studies across diverse health system contexts document that reform packages can raise headline performance while leaving substantial geographic and socioeconomic disparities in service availability and affordability largely intact (Dubas-Jakóbczyk et al., 2020; Korenstein et al., 2016). This divergence is particularly pronounced when reforms shift administrative burden to users, restructure referral pathways, or alter cost-sharing arrangements in ways that increase out-of-pocket exposure for lower-income households. Evidence from Massachusetts and the Affordable Care Act reforms in the United States illustrates that even comprehensive coverage expansions produced differential gains by income, race, and region (Pande et al., 2011; Gaffney & McCormick, 2017; Thomas et al., 2018). In comparison, comparative analyses of health system reforms in China and Central and Eastern Europe document analogous patterns of aggregate improvement alongside persistent sub-national and socioeconomic heterogeneity (Chen et al., 2020; Liu et al., 2017; Yan et al., 2021; Dubas-Jakóbczyk et al., 2020).

In post-Soviet and transitional systems specifically, the expansion of formal financing and infrastructure has frequently coexisted with high household out-of-pocket expenditure and persistent informal payment practices (Bukatov & Gimranova, 2023). Structural constraints, including the geographic concentration of specialist services in urban centers, uneven distribution of diagnostic infrastructure, and transport barriers in rural and remote areas, mean that improvements in system-level resource inputs may yield limited gains in perceived access among disadvantaged populations. This body of work implies that reform effectiveness should be assessed not only through average improvements in official indicators but through distributional gradients in access and financial burden across regions and population groups.

1.3. Patient satisfaction, perceived quality, and the performance-perception gap

Patient satisfaction and perceived quality of care are widely used as user-centered indicators of public service performance because they capture the felt consequences of governance design (responsiveness, communication, transparency, and equity) compared with aggregate administrative metrics (Abidova et al., 2021). International evidence shows that subjective evaluations can remain persistently negative even amid measurable improvements in clinical or administrative outcomes, particularly when users routinely encounter long waiting times, discontinuity of care, opaque service entitlements, and weak communication (Levine et al., 2019; Nuri et al., 2019). These patterns were further amplified during the COVID-19 period, when perceived service quality diverged across systems in ways linked to administrative resilience and the quality of public communication rather than to technical outcomes alone (Tuczyńska et al., 2022).

Critically, satisfaction is a multi-determined construct that systematically reflects service-level factors, particularly timeliness, interpersonal quality, and perceived fairness, rather than macro-level performance narratives (Crow et al., 2002). This implies that a measurable discrepancy between official assessments and citizen evaluations is not anomalous but structurally predictable under

conditions where governance reforms improve reporting and aggregate indicators without proportionate gains in the service encounters that shape day-to-day experience.

Kazakhstan represents an analytically important case within this literature. The country has pursued successive modernization programs targeting financing, infrastructure, and governance under centralized administrative steering (Government of the Republic of Kazakhstan, 2019, 2024). Empirical studies document progress in population health outcomes alongside persistent structural frictions, including substantial regional heterogeneity in service experience and continuing household-level financial exposure (Aymagambetov et al., 2020; Bukatov & Gimranova, 2023; Mukayev et al., 2023). Comparative evidence from other post-Soviet systems reinforces the conclusion that central steering may support uniform policy rollout while simultaneously constraining local feedback loops and responsiveness, thereby amplifying the formal improvement and experience deficit (Grazhevskaya et al., 2019). Existing research in this context, however, has predominantly examined administrative indicators and citizen perceptions in isolation rather than operationalizing them into a single evaluative framework that quantifies the performance–perception gap across regions and links that gap to accessibility and satisfaction gradients.

In summary, the reviewed literature converges on three conclusions. Administrative performance metrics and citizen-reported outcomes constitute distinct but complementary dimensions of reform effectiveness, and their alignment cannot be assumed but must be empirically verified. Distributional inequalities in access and affordability are not resolved automatically by coverage expansion or infrastructure investment, particularly in contexts with pronounced regional and socioeconomic stratification. Patient satisfaction and perceived quality are governance-relevant indicators driven primarily by service-level responsiveness (notably communication, transparency, and timeliness) rather than by official performance narratives. These conclusions collectively support the analytical design of the present study and provide the empirical and theoretical grounding for each of its three hypotheses.

This study aims to quantify the divergence between region-level official assessments of health-care reform effectiveness and citizen-reported service experience in Kazakhstan over the period 2019–2024, and to examine whether this divergence is systematically patterned by inequalities in access and affordability across regions and sociodemographic groups.

Based on the reviewed literature, three testable hypotheses are advanced.

- H1: *Regions with stronger administrative performance are associated with higher citizen-perceived service quality and satisfaction.*
- H2: *Financing and infrastructure measures are associated with higher perceived accessibility and lower affordability constraints across regions and sociodemographic groups.*
- H3: *A systematic performance–perception gap exists such that official assessments of reform effectiveness are more favorable than citizen evaluations.*

1.4. Conceptual model

Based on the reviewed literature and the three hypotheses, the study adopts an integrated conceptual model linking administrative performance, citizen-reported experience, and access-related barriers. This model is presented in Figure 1.

The analytical framework integrates two complementary measurement levels. At the regional level, an official effectiveness index is constructed from standardized administrative indicators (life expectancy dynamics, cause-specific mortality reductions, vaccination coverage, and infrastructure investment), representing the administrative dimension of reform performance. At the individual level, a public perception index is derived from survey-based measures of perceived service quality, overall satisfaction, and perceived reform effectiveness, representing the citizen-experienced dimension. The performance–perception gap score is operationalized as the normalized difference between these two indices for each respondent's region, with higher values indicating greater divergence in favor of official assessments. Accessibility constraints and

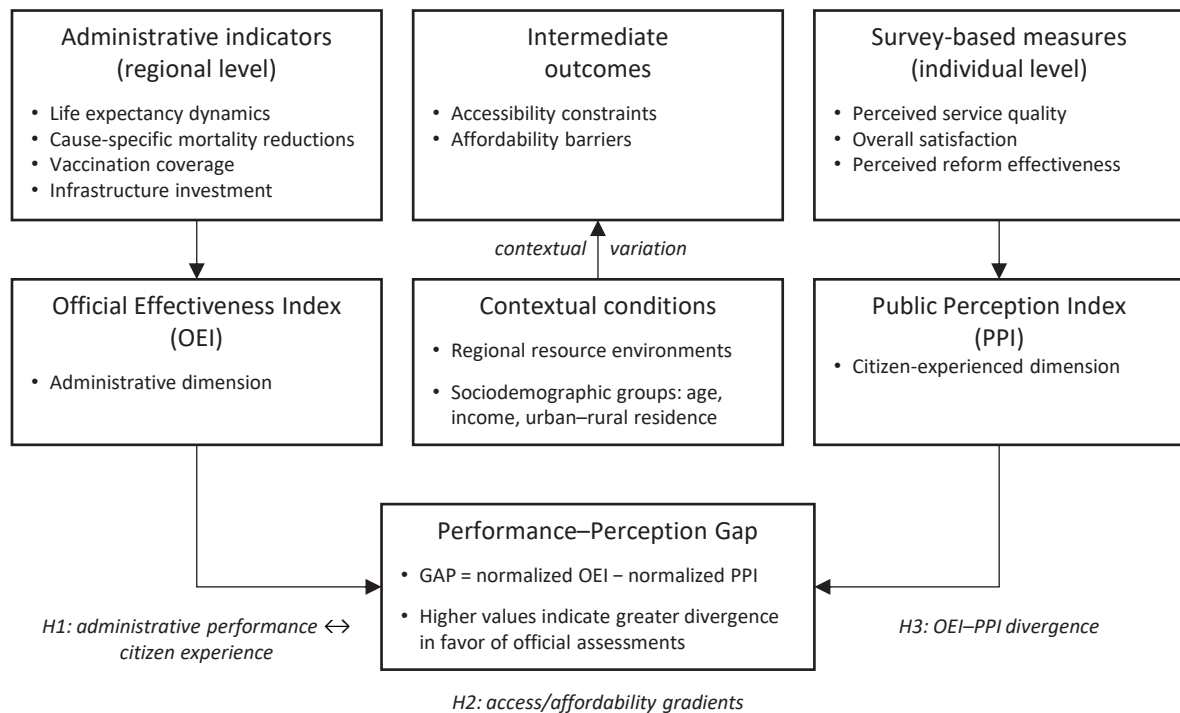


Figure 1. Conceptual framework linking administrative reform performance, citizen experience, and the performance–perception gap

affordability barriers are positioned as intermediate outcomes that mediate the relationship between administrative inputs and citizen-reported experience and are themselves hypothesized to vary systematically across regional resource environments and sociodemographic groups. This framework enables the simultaneous testing of all three hypotheses within a single integrated evaluative design and provides a basis for identifying specific governance bottlenecks (access barriers, transparency deficits, and perceived inequity) in regions where administrative progress is least reflected in citizen experience.

2. METHODS

This study used a mixed-methods design with a cross-sectional comparative framework to evaluate healthcare reforms in Kazakhstan during 2019–2024. Survey-based evidence from a nationwide population sample was integrated with regionally disaggregated official health statistics (2019–2023) and documentary analysis of national policy and legislative materials (2019–2024) governing system organization, financing, benefits, primary care, and digitalization. Integration was implemented by constructing stan-

dardized region-level indices from official indicators and linking them to citizen-reported outcomes to estimate performance–perception gap scores. The design was specified to test three hypotheses concerning perceived service quality and satisfaction (*H1*), perceived accessibility and financial barriers (*H2*), and divergence between official reform-effectiveness assessments and citizen evaluations (*H3*).

2.1. Primary data (population survey)

Primary data were collected via face-to-face structured interviews with adults ($n = 1,940$) across all 20 regions of Kazakhstan (approximately 97 respondents per region). The sampling strategy used stratification by region and urban/rural residence; within strata, households were selected using a random-route procedure, and one eligible respondent was selected using within-household random selection. Eligibility criteria were age ≥ 18 years; residence in the region for ≥ 12 months; and at least one contact with public or private healthcare services within the preceding 24 months. We designed, planned, and supervised the fieldwork, bearing full academic responsibility for the research design, data collection protocol, and analytical decisions. Trained interviewers conducted

data collection from March to August 2024 under the direct supervision of the research team at Turan University. We were responsible for all stages of instrument development, interviewer training, fieldwork coordination, and quality control. This study therefore constitutes our own original empirical research. Kazakhstan was selected as the study setting because it represents a prototypical post-Soviet transitional health system that has implemented successive reform packages under centralized administrative governance while exhibiting pronounced regional disparities in healthcare infrastructure, financing, and population health outcomes; this configuration makes it a uniquely informative case for studying the performance–perception gap at the subnational level. The 2024 data-collection period was chosen deliberately: it falls within the final phase of the State Health Development Program 2020–2025 and the implementation horizon of the Concept for Healthcare Development until 2026, thereby capturing citizen experience at a point when reform effects on service delivery should, in principle, be observable and measurable at the point of care. Face-to-face structured interviewing was employed rather than self-administered or online modes because the target population spans all 20 regions, including rural and remote areas with heterogeneous digital literacy and infrastructure, and because in-person administration minimizes item nonresponse and measurement error in attitudinal survey instruments. All interviews were conducted in the respondent's preferred language (Kazakh or Russian) to ensure comprehension and reduce response bias. Table 1 summarizes respondent characteristics.

2.2. Secondary data (official statistics)

Official health statistics for 2019–2023 (including life expectancy and selected indicators such as cause-specific mortality and vaccination coverage) were obtained from the Ministry of Health and official statistical publications, disaggregated by region where available.

Documentary analysis covered nationally applicable policy programs and legal/regulatory acts directly relevant to healthcare organization, access, quality, financing arrangements, primary care, benefits, and digitalization. Core documents included the State Health Development Program 2020–2025 and the Concept for Healthcare Development until 2026, complemented by relevant legislative and regulatory acts issued within 2019–2024.

The questionnaire comprised 65 items covering socio-demographics, healthcare utilization, perceived service quality, overall satisfaction, accessibility constraints, affordability barriers, and perceived reform effectiveness. Most attitudinal measures used five-point Likert response formats (1 = very low; 5 = very high). The instrument was administered in Kazakh and Russian and underwent translation and back-translation procedures; it was pilot-tested prior to implementation. Multi-item subscales demonstrated acceptable internal consistency (Cronbach's $\alpha > 0.78$). The full 65-item instrument, organized according to these thematic blocks, is provided in Appendix A. The content and structure of the questionnaire were determined by the three study hypotheses.

Table 1. Respondent characteristics (n = 1,940)

Characteristic	Category	n	%
Age group (years)	18–25	454	23.4
	26–30	421	21.7
	31–40	384	19.8
	41–50	328	16.9
	51–60	217	11.2
	61+	136	7.0
Gender	Female	1,016	52.4
	Male	924	47.6
Place of residence	Urban	1,164	60.0
	Rural	776	40.0

Note: Urban/rural stratification was applied with 60% urban and 40% rural respondents, reflecting Kazakhstan's demographic distribution. Within each of the 20 regions, approximately 58 respondents were sampled from urban areas and 39 from rural areas, ensuring regional representation in both strata.

Sociodemographic and utilization items (blocks 1 and 2) were included to enable subgroup analysis by age, income, gender, and urban/rural residence, which are the key stratifiers for *H2*. Perceived service quality and overall satisfaction items (block 3) directly operationalize the citizen-side outcome required for *H1* and *H3*. Accessibility constraint items (block 4) address geographic and organizational barriers to specialized care, constituting the primary *H2* outcome. Affordability barrier items (block 5) capture out-of-pocket payment burden and financial hardship in seeking care, addressing the second *H2* dimension. Perceived reform effectiveness items (block 6) form the individual-level component of the performance-perception gap score required for *H3*. The limitation of the instrument to these six thematic blocks reflects the need to maintain a manageable interview duration (approximately 40 minutes) while ensuring adequate item coverage for multi-item scale construction. Survey results derive exclusively from the 2024 data collection; these data have not been used as the basis of any prior publication.

To align measurement with the hypotheses, three sets of variables were defined. For *H1*, outcomes were overall satisfaction and perceived service quality (including communication-related items), examined alongside region-level official performance indicators (e.g., life expectancy and mortality trends). For *H2*, outcomes included perceived accessibility of specialized services and self-reported affordability barriers, analyzed across regions and socio-demographic groups (including income and urban–rural residence). For *H3*, two composite indices were constructed: a region-level official effectiveness index based on standardized official indicators (e.g., improvements in life expectancy, mortality reductions, infrastructure-related progress, and vaccination coverage) and an individual-level public perception index derived from perceived reform effectiveness, perceived service quality, and overall satisfaction. The perception-gap score was operationalized as the difference between the normalized official effectiveness score for the respondent's region and the respondent's normalized perception score.

2.3. Index construction

The Official Effectiveness Index (OEI_{*r*}) for each region *r* was constructed from four standardized administrative indicators measured over 2019–2023.

The four components are:

- (1) life expectancy dynamics (ΔLE_r), calculated as the change in life expectancy at birth between 2019 and 2023;
- (2) cause-specific mortality reduction ($\Delta Mort_r$), operationalized as the average percentage reduction in cardiovascular and overall mortality rates over the same period;
- (3) vaccination coverage rate ($Vacc_r$), measured as the regional percentage of the target population fully vaccinated in 2023; and
- (4) infrastructure investment index ($Infra_r$), expressed as per capita healthcare capital expenditure averaged over 2019–2023 (in constant 2023 KZT).

Each indicator was min-max normalized to a 0–100 scale across all 20 regions using the formula:

$$X'_r = \frac{(X_r - X_{\min})}{(X_{\max} - X_{\min})} \cdot 100, \quad (1)$$

where higher values indicate better performance. The four normalized components were aggregated with equal weights (0.25 each), as no theoretically grounded basis for differential weighting was available:

$$OEI_r = 0.25 \cdot \Delta LE'_r + 0.25 \cdot \Delta Mort'_r + 0.25 \cdot Vacc'_r + 0.25 \cdot Infra'_r. \quad (2)$$

The resulting OEI_{*r*} scores range from 0 to 100, with one unique value per region ($n = 20$ regional observations). All raw indicator values were sourced from official Ministry of Health statistical publications disaggregated by region.

The Public Perception Index (PPI_{*i*}) was constructed at the individual level ($i = 1, \dots, 1940$) from three survey-based composite subscales, each derived from the multi-item Likert blocks described in the questionnaire. The three components are: (1) perceived service quality (PSQ_{*i*}), the mean score of items Q9–Q22 (Block 3; $\alpha = 0.83$); (2) overall satisfaction (Sat_{*i*}), measured by item Q10 as a single-item global indicator; and (3) perceived reform

effectiveness (PRE_i), the mean score of items Q51–Q65 (Block 6; $\alpha = 0.86$). All subscale scores were rescaled from the original 1–5 Likert range to a 0–100 scale using the formula:

$$S' = \frac{(S - 1)}{4} \cdot 100. \quad (3)$$

The three rescaled components were then aggregated with equal weights (0.33 each):

$$PPI_i = 0.33 \cdot PSQ'_i + 0.33 \cdot Sat'_i + 0.33 \cdot PRE'_i. \quad (4)$$

The Performance–Perception Gap Score (GAP) was then computed as the signed difference between the official and citizen-side scores:

$$GAP_i = OEI_{r(i)} - PPI_i, \quad (5)$$

where $r(i)$ denotes the region of respondent i . Positive values indicate that the official assessment is more favorable than the citizen evaluation; negative values indicate the reverse. Both OEI and PPI are expressed on the same 0–100 scale, ensuring that gap scores are directly interpretable in substantive terms.

Data quality checks included distributional screening, internal-consistency assessment for multi-item measures, and outlier verification. Item nonresponse was minimal; there were no missing values on the key variables used in the main models (region, gender, major outcomes, and core covariates). For composite indices, occasional missing items were handled by within-scale mean substitution when required for multivariable models; descriptive analyses used available cases. Robustness checks using resampling procedures for selected estimates supported the stability of the primary parametric results.

For all models in which a region-level predictor (i.e., OEI) was entered as a predictor of individual-level outcomes, standard errors were computed using cluster-robust estimation at the regional level (20 clusters), thereby accounting for the non-independence of respondents within the same region. Using cluster-robust standard errors, rather than treating regional observations as independent, preserves the validity of infer-

ence in cross-level models without requiring full multilevel modeling; this approach is consistent with standard practice when cluster sizes are approximately equal and the number of clusters is sufficient (Cameron & Miller, 2015). However, it is important to acknowledge a key methodological constraint: the official effectiveness index (OEI) contains only 20 unique values – one per region – and all cross-level analyses that use OEI as a predictor of individual-level outcomes are therefore based on a between-cluster design with $n = 20$ effective independent observations at the region level. Cluster-robust standard errors partially address this by accounting for within-cluster correlation, but they do not fully resolve the issue of limited between-cluster degrees of freedom. For *H3* (the paired comparison of official and citizen indices) and for the region-year analysis of the gap described below, analyses were therefore also conducted at the region level ($n = 20$ regional means) to ensure that inferential statistics reflect the actual number of independent regional units. Because the Public Perception Index was collected in a single wave (2024), the region-year analysis holds each region's 2024 PPI fixed and lets only the official effectiveness index vary across 2020–2023; it is consequently reported below as a benchmarked trend in official performance relative to 2024 perception levels, not as a measured trend in public perception. As an additional robustness check, wild cluster bootstrap resampling (MacKinnon & Webb, 2018) was applied to cross-level regression estimates to verify that p -values remain below the significance threshold given the small number of clusters. Where the Pearson correlations reported in Table 2 involve a region-level predictor paired with individual-level outcomes, these should be interpreted as ecological-level associations ($n = 20$) rather than individual-level estimates; when recomputed at this correct sample size, these correlations are no longer statistically significant (Table 2), unlike the originally reported individual-level figures.

Given the equal allocation of respondents across regions and explicit stratification by urban/rural status, analyses were conducted on the unweighted sample and are interpreted primarily as comparative (regional and subgroup) estimates rather than as nationally weighted prevalence estimates.

The study followed the principles of the Declaration of Helsinki and relevant national regulations for research involving human participants. Respondents were informed about the study objectives, voluntary participation, and confidentiality and provided verbal informed consent before the interview. No personally identifiable information was retained in the analytical dataset. Official statistics and policy documents were analyzed in aggregated form and derived from publicly available sources. Participation was entirely voluntary, and respondents were explicitly informed of their right to decline or discontinue at any stage without consequence. To ensure impartiality, interviewers received standardized training and were required to follow a fixed script without prompting or leading respondents. Anonymity was preserved throughout: no names, addresses, or other direct identifiers were recorded on survey forms; respondents were identified only by a region and stratum code. The research protocol was reviewed and approved by the Research Ethics Committee of Turan University. Participation was not incentivized, and the voluntary nature of the survey was reconfirmed verbally at the outset of each interview.

3. RESULTS

The analyses draw on the nationwide population survey described in the Methods ($n = 1,940$; 20 regions). Results are reported by hypothesis. To make the core predictions of *H1* and *H2* explicit, we first report direct association tests linking region-level official measures to citizen-reported outcomes; these tests are summarized in Table 2. We then provide descriptive distributions and subgroup analyses to clarify where administrative performance aligns – or diverges – from lived service experience.

3.1. Healthcare quality improvements and patient satisfaction (*H1*)

Official indicators improved between 2020 and 2023: life expectancy rose from 73.2 to 74.2 years (+1.4%), infant mortality declined from 8.4 to 7.1 per 1,000 live births (–15.5%), cardiovascular mortality fell by 6.9%, and overall mortality declined by 5.7%. In the survey, perceived service quality was rated as high/very high by 28.35% of respondents, average by 36.08%, and low/very low by 35.56% (Figure 1). Mean satisfaction was near the

Table 2. Direct tests of hypotheses *H1* and *H2*

Hypothesis	Independent Variable	Dependent Variable	Test	β / r	p	N	Interpretation
<i>H1</i> Core Test	Official regional performance index*	Perceived service quality	Linear regression	$\beta = 0.18$	0.041	1,940	Positive but weak association
<i>H1</i> Core Test	Official regional performance index*	Overall satisfaction	Linear regression	$\beta = 0.14$	0.089	1,940	Positive but not significant
<i>H1</i> Correlation	Official regional performance index*	Perceived service quality	Pearson correlation	$r = 0.21$ (regional means, 95% CI [-0.26, 0.60])	0.374	20	Weak association, not significant at regional level ($n = 20$)
<i>H2</i> Core Test	Regional healthcare financing per capita**	Perceived accessibility of specialized services	Linear regression	$\beta = 0.29$	0.003	1,940	Positive significant association
<i>H2</i> Core Test	Regional infrastructure investment index***	Affordability barriers (reversed)	Linear regression	$\beta = 0.22$	0.017	1,940	Positive significant association
<i>H2</i> Correlation	Regional healthcare financing per capita**	Perceived accessibility	Pearson correlation	$r = 0.34$ (regional means, 95% CI [-0.12, 0.68])	0.142	20	Moderate association, not significant at regional level ($n = 20$)

Note: β denotes standardized regression coefficients; r denotes Pearson correlations. Regressions use cluster-robust standard errors at the regional level (20 clusters). For the two Pearson correlations involving the region-level official performance index (*H1* Correlation and *H2* Correlation rows), the table now reports the sample size, correlation coefficient, p -value, and 95% confidence interval computed directly from the $n = 20$ regional means, rather than from the 1,940 individual respondents. At this correct sample size, neither region-level correlation reaches conventional statistical significance ($p = 0.374$ and $p = 0.142$, respectively), and both confidence intervals include zero. This differs from the earlier individual-level figures ($r = 0.21$, $p = 0.028$; $r = 0.34$, $p < 0.001$), which had incorrectly treated the 1,940 respondents as independent replicates of a variable that in fact varies only at the level of the 20 regions. The corrected, non-significant regional-level results are consistent with the modest, non-significant regression coefficient already reported for the *H1* association ($\beta = 0.18$, $p = 0.041$); the *H2* regression coefficients are estimated at the individual level with cluster-robust standard errors and are unaffected by this correction.

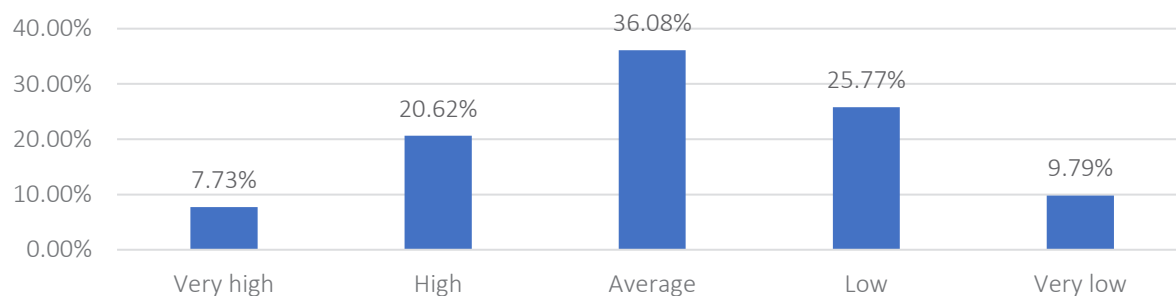


Figure 2. Distribution of respondents' answers to the question "How do you assess the overall quality of medical services in Kazakhstan?"

neutral midpoint ($M = 2.97$, $SD = 1.23$). Consistent with the direct test in Table 2, the official regional performance index shows a positive but weak association with perceived service quality, while the association with overall satisfaction is not statistically significant at the 5% level.

Figure 2 displays the distribution of perceived service quality ratings. Age-group differences were statistically significant ($F(5, 1934) = 16.13$, $p < 0.001$, $\eta^2 = 0.04$): respondents aged 18 to 25 reported higher ratings ($M = 3.21$, $SD = 1.18$) than those aged 61+ ($M = 2.45$, $SD = 1.31$). Post-hoc Tukey tests indicated statistically significant differences between age groups ($p < 0.05$).

Table 3 reports the same quality assessments by age group. Positive assessments (very high + high) declined from 34.88% among respondents aged 18 to 25 to 18.01% among those aged 61+, while negative assessments (low + very low) increased from 25.26% to 60.06%.

Table 3. Distribution of responses to "How do you assess the overall quality of medical services in Kazakhstan?" by age group

Age group	Very high	High	Average	Low	Very low
18-25	10.32%	24.56%	39.86%	20.28%	4.98%
26-30	8.91%	21.29%	39.60%	20.30%	9.90%
31-40	7.43%	18.47%	35.01%	24.94%	14.15%
41-50	6.72%	17.09%	36.13%	24.93%	15.13%
51-60	5.74%	13.85%	30.41%	25.00%	25.00%
61+	4.96%	13.05%	21.93%	30.03%	30.03%

Multiple regression showed that waiting times ($\beta = -0.342$, $p < 0.001$), staff communication ($\beta = 0.289$, $p < 0.001$), and facility conditions ($\beta = 0.198$, $p < 0.001$) were statistically significant pre-

dictors of patient satisfaction, together explaining 34.7% of its variance ($R^2 = 0.347$, $F(3, 1936) = 342.19$, $p < 0.001$).

3.2. Healthcare accessibility and infrastructure development (H2)

Consistent with the direct tests reported in Table 2, regional resource measures are positively associated with citizen-reported access outcomes. Higher regional healthcare financing per capita is linked to higher perceived accessibility of specialized services ($\beta = 0.29$, $p = 0.003$), and the infrastructure investment index is linked to lower affordability constraints (reverse-coded affordability barriers, $\beta = 0.22$, $p = 0.017$). These associations are complemented below by distributional evidence and subgroup contrasts that indicate where access remains constrained.

Perceived accessibility of specialized services remains limited for most respondents. Only 10.36% rated specialized services as very/easily accessible, 35.52% as moderately accessible, and 54.12% as difficult or impossible to access. Accessibility differed markedly across regions (one-way ANOVA: $F(19, 1920) = 59.61$, $p < 0.001$, $\eta^2 = 0.37$), with the highest mean accessibility in major cities (e.g., Astana $M = 3.78$; Almaty $M = 3.72$; Shymkent $M = 3.21$) and the lowest in selected regions (e.g., Mangistau $M = 1.98$; Kyzylorda $M = 2.09$). An urban-rural comparison confirmed a large disparity ($t(1938) = 23.52$, $p < 0.001$, $d = 1.24$), with higher accessibility in urban areas ($M = 3.45$) than in rural areas ($M = 2.31$).

Perceived infrastructure-related improvements were also reported, but they do not eliminate ac-

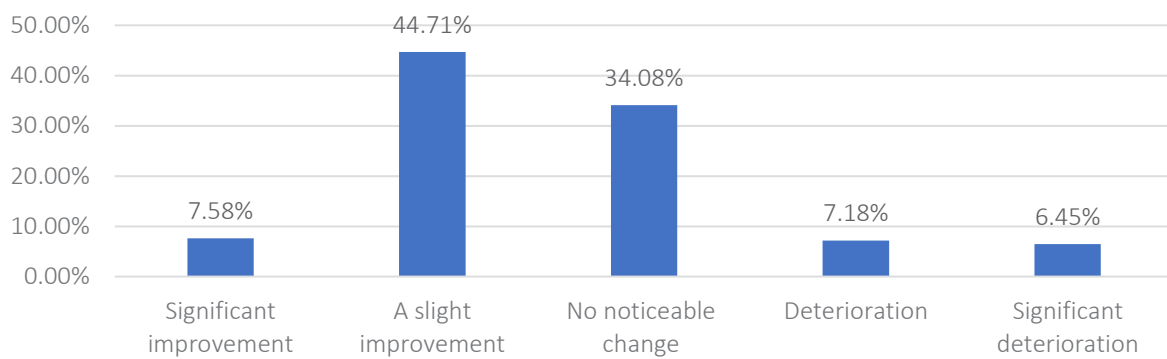


Figure 3. Distribution of respondents' answers to the question "Do you feel that the conditions of healthcare facilities have improved in recent years"

cess constraints. As shown in Figure 3, 52.29% of respondents reported that facility conditions had improved in recent years (7.58% "significant improvement" and 44.71% "slight improvement"), while 34.08% reported no noticeable change and 13.63% reported deterioration (7.18% "deterioration" and 6.45% "significant deterioration").

Affordability constraints remain socially patterned. Cost-related barriers differed significantly by socioeconomic status: income was associated with affordability constraints ($\chi^2(12) = 106.84$, $p < 0.001$, Cramér's $V = 0.23$), with 67.8% of low-income respondents versus 23.4% of high-income respondents identifying high cost as a main barrier. Taken together, the results support *H2*: financing and infrastructure measures predict better perceived accessibility and lower affordability constraints, yet substantial regional and socio-demographic gradients persist.

3.3. Discrepancies between official reports and public perception (*H3*)

To test *H3*, we compared a standardized region-level official effectiveness index (assigned to each respondent based on their region) with an individual-level public perception index and calculated a performance-perception gap score as the difference between the normalized official score and the respondent's normalized perception score (higher values indicate a more favorable official assessment relative to citizen evaluation).

The composite official-effectiveness score exceeded the composite public-perception score ($M =$

73.4, $SD = 8.7$ vs. $M = 58.2$, $SD = 12.3$), yielding a large, systematic divergence ($d = 1.38$). Because the OEI takes only 20 unique regional values, the appropriate primary test for *H3* is a paired *t*-test conducted at the regional level ($n = 20$ region-mean pairs), comparing mean OEI with mean PPI aggregated per region: $t(19) = 14.87$, $p < 0.001$, $d = 1.38$. The corresponding individual-level paired *t*-test ($t(1939) = 29.74$, $p < 0.001$) is reported for descriptive completeness but should not be interpreted as providing 1,940 independent degrees of freedom, since all respondents within the same region share an identical OEI value. Both the region-level and individual-level tests point to the same substantive conclusion: there is a systematic tendency for official reform-effectiveness assessments to be more favorable than citizen evaluations.

The magnitude of the perception gap varied across demographic and territorial contexts. A two-way ANOVA on the perception-gap score indicated significant effects of age ($F(5, 1925) = 12.87$, $p < 0.001$, $\eta^2 = 0.03$) and region ($F(19, 1925) = 16.87$, $p < 0.001$, $\eta^2 = 0.14$), as well as a significant age-region interaction ($F(95, 1925) = 1.46$, $p < 0.001$, $\eta^2 = 0.07$). Thus, divergence between official and citizen assessments is not uniform across the country and population groups.

In multivariable models, predictors jointly explained 31.8% of the variance in the perception-gap score ($R^2 = 0.318$, $F(4, 1935) = 225.51$, $p < 0.001$). A regional analysis of the gap over the 2020–2023 period requires an important clarification of what was actually measured. The Public Perception Index (PPI) is a survey-based measure collected in a single cross-sectional wave in 2024; no repeat-

ed perception measurements exist for 2020, 2021, 2022, or 2023. The reported change in the mean regional gap, from 8.7 to 15.3 points, was obtained by holding each region's 2024 PPI value fixed as a static perception benchmark and recalculating the region-level Official Effectiveness Index (OEI) at each of the four annual time points using the underlying official statistics (life expectancy, mortality, vaccination coverage, and infrastructure investment), which are available on a yearly basis for 2020–2023. The apparent widening of the gap therefore reflects year-on-year change in the official indicators evaluated against one constant perception reference point, not a documented change in public perception itself. Because only the numerator (OEI) varies across the four time points while the denominator (PPI) is fixed by construction, this analysis cannot be interpreted as evidence of a genuine time trend in public perception. The terminology used elsewhere in this study has been revised accordingly to refer to it as a “benchmarked official-performance trend relative to 2024 perception levels” rather than a “time trend in the performance–perception gap.” The region-year regression ($n = 20$ regions $\times$ 4 time points = 80 observations; $F(1, 78) = 92.16$, $p < 0.001$, $R^2 = 0.54$) remains statistically informative about the trajectory of the official indicators, but it does not by itself demonstrate that citizens' evaluations became more negative, or more positive, over 2020–2023, since perception was measured only once. The effective number of independent regional units in this model is 20, and the high F -statistic should be interpreted with this repeated-measures, single-perception-wave structure in mind.

Effect-size estimates indicate that the cross-sectional divergence between official and public assessments is substantively large rather than merely statistically detectable: the difference between official and public indices corresponds to $d = 1.38$. The region-year model of official-indicator change against the fixed 2024 perception benchmark yields a large f^2 ($f^2 = 1.174$, $R^2 = 0.54$). This indicates that year-on-year movement in the official effectiveness index accounts for a substantial share of variance in the benchmarked gap. However, this statistic describes the trajectory of the administrative indicators alone and should not be read as evidence of strengthening explanatory power over changes in public perception, which was not repeatedly measured.

4. DISCUSSION

This study examined whether Kazakhstan's recent healthcare reforms are associated with (*H1*) higher perceived service quality and satisfaction, (*H2*) broader and more equitable access and affordability, and (*H3*) closer alignment between administrative performance metrics and citizens' evaluations. Overall, the findings indicate a persistent divergence between improvements captured by official indicators and the population's reported experience of care, with substantial heterogeneity across regions and sociodemographic groups.

4.1. *H1* (quality and satisfaction)

Official statistics for 2020–2023 suggest continued improvement in population-level outcomes (life expectancy +1.4%; infant mortality –15.5%; cardiovascular and overall mortality –6.9% and –5.7%). However, citizen evaluations remain moderate, and satisfaction is lower in older age groups. Importantly, the direct hypothesis test indicates only modest alignment between administrative performance and perceived service quality (Table 2: $\beta = 0.18$, $p = 0.041$), while the association with overall satisfaction is weaker and not statistically significant (Table 2: $\beta = 0.14$, $p = 0.089$). This pattern implies that aggregate performance gains translate only partially into perceived improvements at the point of care. Our results corroborate the “performance-satisfaction gap” identified in broader health management literature, suggesting that macro-level indicators do not always reflect individual patient experiences. Within individual experience, regression results show that waiting times, staff communication, and facility conditions are the strongest correlates of satisfaction. This finding aligns directly with the evidence from Abidova et al. (2021) and Levine et al. (2019), confirming that patient experience is primarily shaped by timeliness and interpersonal processes rather than by macro-level performance summaries. Furthermore, our data support the observations of Tuczyńska et al. (2022), illustrating how technical progress can coexist with mixed assessments of service delivery during periods of systemic pressure.

4.2. *H2* (accessibility and affordability)

Evidence supports *H2*, but with clear distributional constraints. Respondents report improvement in facility conditions, which provides empirical support for the success of national modernization priorities outlined by the Government of the Republic of Kazakhstan (2019, 2024) and documented by Baizhiyenova et al. (2023). At the same time, access to specialized care and affordability remain stratified by region, urban–rural residence, and income. The direct tests show that regional health-care financing per capita is positively associated with perceived accessibility of specialized services (Table 2: $\beta = 0.29$, $p = 0.003$), and that infrastructure investment is associated with lower affordability constraints (Table 2: $\beta = 0.22$, $p = 0.017$). Yet, the persistence of strong regional and socioeconomic gradients indicates that resource inputs, while beneficial on average, do not automatically eliminate geographic and household-level barriers. Our findings align with patterns documented in other international contexts, where coverage expansion and financing reforms improved access in aggregate but did not remove geographic and socioeconomic disparities (Pande et al., 2011; Miller, 2012; Gaffney & McCormick, 2017; Thomas et al., 2018). Moreover, the disparities observed in our study mirror the heterogeneous reform outcomes reported in China and Central/Eastern Europe (Chen et al., 2020; Liu et al., 2017; Yan et al., 2021; Dubas-Jakóbczyk et al., 2020). The observed income gradient in cost barriers further reinforces the evidence from Bukatov and Gimranova (2023) that household financial exposure remains policy-salient in Kazakhstan.

4.3. *H3* (performance–perception gap)

H3 is strongly supported at the cross-sectional level. Composite official effectiveness metrics exceed public evaluations (2024), and official indicators moved further away from the fixed 2024 perception benchmark between 2020 and 2023. This systematic divergence provides a specific national illustration of the broader theoretical concern that reforms may yield limited value relative to policy narratives (Korenstein et al., 2016). Our findings also resonate with Farley's (2009) perspective, suggesting that perceived impact remains constrained when broader social determinants are overlooked. Finally, by establishing that personal experience outweighs me-

dia exposure, our study extends the evidence of Aymagambetov et al. (2020) and Zharlygassinov et al. (2023), demonstrating that alignment between administrative assessments and public evaluations is fundamentally shaped by direct encounters with the system. Because public perception was measured only once, in 2024, this study cannot establish whether citizen evaluations themselves became more negative over 2020–2023; it can only show that official performance indicators diverged increasingly from present-day (2024) perception levels over that period.

4.4. Limitations

Several limitations should be noted. The design is cross-sectional, which limits causal inference; key outcomes are perception-based (including retrospective assessments); the stratified sample is unweighted and interpreted primarily in comparative terms; and official indicators are linked at an aggregate regional level through composite indices. A further methodological limitation concerns statistical inference in cross-level analyses: the official effectiveness index contains only 20 unique regional values. This means that all analyses linking the OEI to individual-level outcomes (including the Pearson correlations and the paired *t*-test for *H3*) are fundamentally based on between-region variation across $n = 20$ clusters rather than $n = 1,940$ individuals. Although cluster-robust standard errors were applied to the regression models, this correction is not fully sufficient for the paired *t*-test and Pearson correlations. To address this, we conducted the *H3* primary test at the regional level ($n = 20$ region-mean pairs) and recomputed the Pearson correlations in Table 2 using regional aggregates ($n = 20$); at the correct sample size, these two correlations are no longer statistically significant, indicating that the original individual-level figures had overstated inferential precision. The *H3* paired-comparison result, by contrast, remained significant at the regional level and is therefore more robust to this correction. A related, and more severe, limitation concerns the region-year analysis of the performance–perception gap over 2020–2023: because the Public Perception Index was collected in a single cross-sectional wave (2024), this analysis necessarily holds perception fixed at its 2024 level and varies only the official effectiveness index across years. Consequently, the reported widening of the region-year gap ($F(1, 78)$

= 92.16, $p < 0.001$, $R^2 = 0.54$) documents change in the official indicators relative to a static perception benchmark and must not be interpreted as a measured change in public perception over time. Genuinely tracking how public perception evolves would require repeated survey waves, which this cross-sectional design does not provide. Future research should consider applying full multilevel or hierarchical modeling to appropriately partition individual-level and regional-level variance, employing permutation-based inference designed for small numbers of clusters, and, most importantly, implementing repeated perception surveys to allow a genuine longitudinal test of the performance–perception gap. Despite these constraints, the national scope, the internal consistency of multi-item measures, and the convergence of direct tests with subgroup patterns support the robustness of the main conclusions.

4.5. Implications

In summary, Kazakhstan's reforms are accompanied by measurable gains in health outcomes and visible improvements in facility conditions, but these advances are not matched by proportionate improvements in perceived quality, equitable accessibility, or convergence between official and public assessments. Policy implications follow directly from the observed correlates and gradients: reductions in waiting times, improvements in communication and service responsiveness, and targeted measures to address accessibility and affordability constraints in lower-access and lower-income settings. More broadly, the findings support evaluation approaches that integrate user-reported outcomes with administrative performance metrics to diagnose where system improvements are not reaching citizens in practice.

CONCLUSION

This study aimed to quantify the divergence between region-level official assessments of healthcare reform effectiveness and citizen-reported service experience in Kazakhstan over the period 2019–2024, and to examine whether this divergence is systematically patterned by inequalities in access and affordability across regions and sociodemographic groups.

Official administrative statistics document measurable gains over the study period; life expectancy increased by 1.4%, infant mortality declined by 15.5%. Measured against the fixed 2024 public-perception benchmark, the composite performance–perception gap widened from 8.7 to 15.3 points between 2020 and 2023 ($d = 1.38$). Since perception itself was assessed only once, this widening reflects change in official performance indicators rather than a demonstrated change in citizen evaluations. However, citizen assessments remained substantially below the levels that administrative indicators would suggest, with only 28.35% of respondents rating service quality as high (7.73% very high + 20.62% high) and overall satisfaction near the neutral midpoint. Regional financing per capita is positively associated with perceived accessibility ($\beta = 0.29$, $p = 0.003$) and infrastructure investment with lower affordability constraints ($\beta = 0.22$, $p = 0.017$), but pronounced urban-rural and income gradients persist: 54.1% of respondents rated specialized care as difficult or impossible to access, and 67.8% of low-income respondents identified cost as a principal barrier. At the individual level, waiting times, staff communication quality, and facility conditions emerge as the strongest predictors of satisfaction ($R^2 = 0.347$), confirming that the quality of direct service encounters shapes citizen evaluations more powerfully than aggregate policy narratives.

These findings lead to three conclusions. First, administrative performance and citizen-reported experience constitute empirically distinguishable and systematically divergent dimensions of reform effectiveness in Kazakhstan, driven by governance frictions at the point of care that aggregate indicators do not capture. Second, financing and infrastructure improvements, while directionally beneficial, are insufficient to eliminate geographic and socioeconomic inequalities in access and affordability, which remain both large in magnitude and persistent over time. Third, the widening performance–perception gap over 2020–2023 indicates that the current evaluation architecture systematically overestimates citizen-

experienced reform progress, calling for governance interventions that target responsiveness, communication quality, and equitable access – particularly in lower-resource regions – as the 2026 healthcare concept reaches completion.

AUTHOR CONTRIBUTIONS

Conceptualization: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Data curation: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Formal analysis: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Funding acquisition: Dina Razakova.

Investigation: Talgat Zharlygassinov.

Methodology: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Project administration: Talgat Zharlygassinov, Aigerim Sekerbayeva.

Resources: Aigerim Sekerbayeva.

Software: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Supervision: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Validation: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Visualization: Talgat Zharlygassinov.

Writing – original draft: Talgat Zharlygassinov, Dina Razakova, Aigerim Sekerbayeva.

Writing – review & editing: Talgat Zharlygassinov, Aigerim Sekerbayeva.

REFERENCES

1. Abidova, A., da Silva, P. A., & Moreira, S. (2021). The mediating role of patient satisfaction and perceived quality of health-care in the emergency department. *Medicine*, 100(11), Article e25133. <https://doi.org/10.1097/MD.00000000000025133>
2. Alhrahsheh, R., & Ivanova, G. Y. (2022). Exploring digital access as a social determinant of health amongst UAE students. *Journal of Social Studies Education Research*, 13(1), 193-216. Retrieved from <https://jsser.org/index.php/jsser/article/view/3748>
3. Andersen, S. T. (2019). eHealth systems' user experiences from healthcare workers. *Research in Social Sciences and Technology*, 4(1), 51-72. <https://doi.org/10.46303/res-sat.04.01.3>
4. Aymagambetov, Y., Grazhevskaya, N., & Tyngisheva, A. (2020). Estimation the effectiveness of public governance of the health system in the context of sustainable development. *Entrepreneurship and Sustainability Issues*, 7(4), 3309-3320. [https://doi.org/10.9770/jesi.2020.7.4\(48\)](https://doi.org/10.9770/jesi.2020.7.4(48))
5. Baizhiyenova, K. T., Omarova, B. K., & Kalmenova, B. T. (2023). History of healthcare development in independent Kazakhstan. *Bulletin of the Karaganda University. History. Philosophy Series*, 109(1), 49-55. Retrieved from <https://history-philosophy-vestnik.buketov.edu.kz/history-philosophy-vestnik/article/view/552>
6. Bromley, P., & Powell, W. W. (2012). From smoke and mirrors to walking the talk: Decoupling in the contemporary world. *The Academy of Management Annals*, 6(1), 483-530. <https://doi.org/10.5465/19416520.2012.684462>
7. Bukatov, Y., & Gimranova, G. (2023). Government measures to address out-of-pocket health expense in Kazakhstan. *Cogent Economics & Finance*, 11(1), Article 2164409. <https://doi.org/10.1080/23322039.2022.2164409>
8. Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317-372. <https://doi.org/10.3368/jhr.50.2.317>
9. Chen, J., Xu, S., & Gao, J. (2020). The mixed effect of China's new health care reform on health insurance coverage and the efficiency of health service utilisation: A longitudinal approach. *International Journal of Environmental Research and Public Health*, 17(5), Article 1782. <https://doi.org/10.3390/ijerph17051782>
10. Crow, R., Gage, H., Hampson, S., Hart, J., Kimber, A., Storey, L., & Thomas, H. (2002). The measurement of satisfaction with health-care: Implications for practice from a systematic review of the literature. *Health Technology Assessment*, 6(32), 1-244. Retrieved from <https://pubmed.ncbi.nlm.nih.gov/12925269/>
11. Dubas-Jakóbczyk, K., Albrecht, T., Behmane, D., Bryndova, L., Dimova, A., Džakula, A., Habicht, T., Murauskienė, L., Scintee, S. G., Smatana, M., Velkey, Z., & Quentin, W. (2020). Hospital reforms in 11 Central and Eastern European countries between 2008 and 2019: A comparative analysis. *Health Policy*, 124(4), 368-379. <https://doi.org/10.1016/j.healthpol.2020.02.003>
12. Farley, T. A. (2009). Reforming health care or reforming health? *American Journal of Public Health*, 99(4), 588-590. <https://doi.org/10.2105/AJPH.2008.158808>

13. Ferlie, E., & McGivern, G. (2014). Bringing Anglo-governmentality into public management scholarship: The case of evidence-based medicine in UK health care. *Journal of Public Administration Research and Theory*, 24(1), 59-83. <https://doi.org/10.1093/jopart/mut002>
14. Footman, K., Roberts, B., Mills, A., Richardson, E., & McKee, M. (2013). Public satisfaction as a measure of health system performance: A study of nine countries in the former Soviet Union. *Health Policy*, 112(1-2), 62-69. <https://doi.org/10.1016/j.healthpol.2013.03.004>
15. Gaffney, A., & McCormick, D. (2017). The Affordable Care Act: Implications for health-care equity. *The Lancet*, 389, 1442-1452. [https://doi.org/10.1016/S0140-6736\(17\)30786-9](https://doi.org/10.1016/S0140-6736(17)30786-9)
16. Government of the Republic of Kazakhstan. (2019, December 26). *On approval of the State Program for Healthcare Development of the Republic of Kazakhstan for 2020-2025* (Resolution No. 982; expired by Resolution No. 725 of October 12, 2021). (In Russian). Retrieved from <https://adilet.zan.kz/rus/docs/P1900000982>
17. Government of the Republic of Kazakhstan. (2024, February 29). *On amendments to the Resolution of the Government of the Republic of Kazakhstan dated November 24, 2022, No. 945 "On the approval of the Concept for the Development of Healthcare in the Republic of Kazakhstan until 2026."* (In Russian). Retrieved from <https://adilet.zan.kz/rus/docs/P2200000945>
18. Grazhevskaya, N., Aimagambetov, Y., & Tyngisheva, A. (2019). Comparative analysis of operation and development of health systems of Ukraine and Kazakhstan. *Bulletin of Taras Shevchenko National University of Kyiv. Economics*, (202), 13-20. <https://doi.org/10.17721/1728-2667.2019/202-1/2>
19. Korenstein, D., Duan, K., Diaz, M. J., Ahn, R., & Keyhani, S. (2016). Do health care delivery system reforms improve value? The jury is still out. *Medical Care*, 54(1), 55-66. <https://doi.org/10.1097/MLR.0000000000000445>
20. Levine, D. M., Landon, B. E., & Linder, J. A. (2019). Quality and experience of outpatient care in the United States for adults with or without primary care. *JAMA Internal Medicine*, 179(3), 363-372. <https://doi.org/10.1001/jamainternmed.2018.6716>
21. Liu, G. G., Vorthems, S. A., & Hong, X. (2017). China's health reform update. *Annual Review of Public Health*, 38, 431-448. <https://doi.org/10.1146/annurev-publ-health-031816-044247>
22. MacKinnon, J. G., & Webb, M. D. (2018). The wild bootstrap for few (treated) clusters. *The Econometrics Journal*, 21(2), 114-135. <https://doi.org/10.1111/ectj.12107>
23. Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340-363. <https://doi.org/10.1086/226550>
24. Miller, S. (2012). The impact of the Massachusetts health care reform on health care use among children. *American Economic Review*, 102(3), 502-507. <https://doi.org/10.1257/aer.102.3.502>
25. Mukayev, A., Satpayeva, Z., Kagalakova, D., Doskeyeva, G., & de Matos Pedro, E. (2023). Assessment of the population's quality of life in Kazakhstan during COVID-19: The effectiveness of public policy. *Problems and Perspectives in Management*, 21(3), 69-83. [https://doi.org/10.21511/ppm.21\(3\).2023.06](https://doi.org/10.21511/ppm.21(3).2023.06)
26. Newman, S., & Lawler, J. (2009). Managing health care under New Public Management. *Journal of Sociology*, 45(4), 419-432. <https://doi.org/10.1177/1440783309346477>
27. Nuri, N. N., Sarker, M., Ahmed, H. U., Hossain, M. D., Beiersmann, C., & Jahn, A. (2019). Experience and perceived quality of care of patients and their attendants in a specialized mental hospital in Bangladesh. *International Journal of Mental Health Systems*, 13, Article 46. <https://doi.org/10.1186/s13033-019-0303-x>
28. Pande, A. H., Ross-Degnan, D., Zaslavsky, A. M., & Salomon, J. A. (2011). Effects of healthcare reforms on coverage, access, and disparities: Quasi-experimental analysis of evidence from Massachusetts. *American Journal of Preventive Medicine*, 41(1), 1-8. <https://doi.org/10.1016/j.amepre.2011.03.010>
29. Suárez-Brito, P., Vázquez-Parra, J. C., Alonso-Galicia, P. E., Cruz-Sandoval, M., & Malagón-Castro, L. E. (2024). Scientific thinking engineering and health sciences students: Competency analysis in a Mexican university. *Journal of Social Studies Education Research*, 15(2), 29-56. Retrieved from <https://jsser.org/index.php/jsser/article/view/5357>
30. Thomas, K. C., Shartzter, A., Kurth, N. K., & Hall, J. P. (2018). Impact of ACA health reforms for people with mental health conditions. *Psychiatric Services*, 69(2), 231-234. <https://doi.org/10.1176/appi.ps.201700044>
31. Tuczynska, M., Staszewski, R., Matthews-Kozanecka, M., Żok, A., & Baum, E. (2022). Quality of the healthcare services during COVID-19 pandemic in selected European countries. *Frontiers in Public Health*, 10, Article 870314. <https://doi.org/10.3389/fpubh.2022.870314>
32. Vamstad, B. S., & Gordon, M. (2018). Integrating public affairs information strategy with organizational practices in healthcare delivery organizations. *Journal of Public Affairs*, 18(4), Article e1828. <https://doi.org/10.1002/pa.1828>
33. Van de Walle, S. (2018). Explaining citizen satisfaction and dissatisfaction with public services. In E. Ongaro & S. Van Thiel (Eds.), *The Palgrave handbook of public administration and management in Europe* (pp. 227-241). Palgrave Macmillan. https://doi.org/10.1057/978-1-137-55269-3_11
34. van Thiel, S., & Leeuw, F. L. (2002). The performance paradox in the public sector. *Public Performance & Management Review*, 25(3), 267-281. <https://doi.org/10.2307/3381236>
35. Yan, C., Liao, H., Ma, Y., & Wang, J. (2021). The impact of health care reform since 2009 on the efficiency of primary health services: A provincial panel data study in China. *Frontiers in Public Health*, 9, Article 735654. <https://doi.org/10.3389/fpubh.2021.735654>
36. Zharlygassinov, T. M., Kupeshova, S. T., & Nurymova, S. K. (2023). Theoretical foundations of management and functioning of the healthcare system in the Republic of Kazakhstan. *Bulletin of "Turan" University*, (4), 83-95. <https://doi.org/10.46914/1562-2959-2023-1-4-83-95>

APPENDIX A

Table A1. Survey questionnaire

Block 1. Sociodemographic Characteristics
Q1. What is your age? (open numeric)
Q2. What is your sex? (1 = Female; 2 = Male)
Q3. What is your place of residence? (1 = Urban; 2 = Rural)
Q4. What is your highest level of education completed? (1 = Primary or less; 2 = Secondary; 3 = Vocational/technical; 4 = Bachelor's degree; 5 = Master's degree or higher)
Q5. How would you describe your household's monthly income level? (1 = Very low; 2 = Below average; 3 = Average; 4 = Above average; 5 = High)
Block 2. Healthcare Utilization
Q6. How many times did you contact a public or private healthcare facility in the past 24 months? (open numeric)
Q7. What type of facility did you most recently use? (1 = Public polyclinic/primary care; 2 = Public hospital; 3 = Private clinic; 4 = Other)
Q8. How long have you resided in your current region? (1 = Less than 12 months; 2 = 1–3 years; 3 = 4–10 years; 4 = More than 10 years)
Block 3. Perceived Service Quality and Overall Satisfaction (5-point scale: 1 = very low; 5 = very high)
Q9. How would you rate the overall quality of healthcare services you received in your region over the past two years?
Q10. How satisfied are you overall with the healthcare services available in your region?
Q11. How would you rate the quality of communication by healthcare staff (clarity of explanations, attentiveness, respect)?
Q12. How would you rate the responsiveness of healthcare services to your needs (timeliness of receiving care, follow-up)?
Q13. How would you rate the condition and equipment of healthcare facilities you visited?
Q14–Q22. [Additional service quality items covering waiting times, transparency of entitlements, complaint-handling, continuity of care, trust in healthcare providers, perceived fairness of service delivery, digital tools usability, information provision, and privacy] (5-point scale)
Block 4. Accessibility Constraints (5-point scale: 1 = no barrier at all; 5 = very serious barrier)
Q23. To what extent is the physical distance to a healthcare facility a barrier to accessing care for you or your household?
Q24. To what extent are long waiting times a barrier to accessing primary care services?
Q25. To what extent is limited availability of specialized care (e.g., specialist physicians, diagnostic equipment) a barrier in your region?
Q26–Q36. [Additional accessibility items covering: referral complexity, transport availability, language/cultural barriers, appointment scheduling difficulty, awareness of entitlements under GCHP, digital registration barriers, availability of medicines, night/weekend access, access for persons with disabilities, access for elderly, and perceived equity of access] (5-point scale)
Block 5. Affordability Barriers (5-point scale: 1 = no problem at all; 5 = very serious problem)
Q37. To what extent do out-of-pocket payments for medicines represent a financial burden for your household?
Q38. To what extent do out-of-pocket payments for consultations or procedures represent a financial burden for your household?
Q39. In the past 12 months, has your household delayed or foregone necessary medical care due to cost? (1 = Never; 2 = Rarely; 3 = Sometimes; 4 = Often; 5 = Always)
Q40–Q50. [Additional affordability items covering informal payments, catastrophic health expenditure, knowledge of covered services under GCHP, perceived fairness of cost-sharing, financial hardship due to hospitalization, ability to purchase prescribed medicines, insurance coverage adequacy, transport costs to facilities, cost of dental care, cost of rehabilitation services, and household financial resilience following illness] (5-point scale)
Block 6. Perceived Reform Effectiveness (5-point scale: 1 = strongly disagree / deteriorated significantly; 5 = strongly agree / improved significantly)
Q51. Overall, how would you assess changes in the quality of healthcare services in your region over the past five years?
Q52. To what extent do you agree that ongoing health reforms have improved access to healthcare services in your region?
Q53. To what extent do you agree that health reforms have made services more affordable for people with lower incomes?
Q54. How would you evaluate the overall effectiveness of healthcare reform in Kazakhstan over the past five years?
Q55–Q65. [Additional reform effectiveness items covering: perceived improvement in primary care, perceived changes in equity of services, trust in the healthcare system, awareness of the Guaranteed Package of Medical Assistance (GCHP), perceived transparency of healthcare management, media influence on reform perceptions, perceived improvement in digitalization, changes in hospital availability, personal experience as the main source of reform assessment, perceived anti-corruption measures in healthcare, and overall confidence in the national health policy direction] (5-point scale)

Note: This questionnaire was administered face-to-face by trained interviewers in Kazakh or Russian (March–August 2024) across all 20 regions of Kazakhstan (n = 1,940). Participation was voluntary and anonymous. All items used five-point Likert scales (1 = very low / strongly disagree; 5 = very high / strongly agree) unless otherwise indicated. A pilot test was conducted with 40 respondents (not included in the analytical sample) prior to main fieldwork. Items Q14–Q22, Q26–Q36, Q40–Q50, and Q55–Q65 represent additional items within each thematic block; the exact wording is available in the original instrument in Kazakh and Russian. Cronbach's α for multi-item subscales ranged from 0.79 to 0.86 across blocks 3–6.