Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Perception vs. Reality: Assessing Knowledge Gaps in Urinary Stone Management Among Medical Trainees and Junior Physicians

Rayan Alqarni

rayan.alqarni11@gmail.com

Imam Mohammad Ibn Saud Islamic University

Rashad M Nassar

Prince Sultan Military Medical City

Abdulaziz M Aleidan

Imam Mohammad Ibn Saud Islamic University

Suhail Albugami

Prince Sultan Military Medical City

Khaled Almodeth

King Saud Medical City

Khalid Alkharashi

Ministry of the National Guard - Health Affairs

Feras Alharbi

Ad Diriyah Hospital

Wesam Khan

University of Tabuk

Research Article

Keywords: Nephrolithiasis, Calcium oxalate, Uric acid, Clinical competence, Self-assessment, Cross-sectional studies

Posted Date: September 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10706650/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Kidney stones are common in Saudi Arabia. The two main types are handled differently: calcium oxalate (CaOx) stones resist pharmacological dissolution, whereas uric acid (UA) stones can be broken down by alkalinising the urine. Junior clinicians are usually the first to image and triage these patients, yet they may read their own competence inaccurately. Whether such a mismatch exists in Saudi trainees has not, as far as we know, been examined.

Methods

Using a bilingual Arabic-English electronic questionnaire, this cross-sectional survey set the factual knowledge of 407 Riyadh-based participants against their self-perceived knowledge; the sample comprised clinical-year medical students, interns, and general practitioners. Four factual items (rescaled to 0–10) and thirteen self-perceived Likert items (Cronbach's $\alpha = 0.834$) were sorted by modified Bloom's thresholds, and the difference between the two, self-perceived minus factual, was taken to index calibration. Analysis drew on Kruskal-Wallis tests and Spearman correlation, together with multivariable regression.

Results

Good factual knowledge proved far more common than a Good self-rating (41.3% vs 6.6%). Self-perception tracked actual performance only loosely (Spearman $\rho = 0.224$, $p < .001$), and the sample underrated itself (mean knowledge gap − 1.41). Medical students were strongest on the factual items but carried the largest negative gap (− 1.92), while general practitioners scored lowest ($p = .018$). Accuracy reached 82.8% for uric acid dissolvability and 88.4% for CT stone density, yet fell to 42.1% for CaOx non-dissolvability. After adjustment, the 34–40 age band predicted weaker factual knowledge ($\beta = -3.63$, $p < .001$), whereas it was cadre that predicted calibration, with interns and general practitioners less under-confident than students ($p = .003$ and .027).

Conclusions

For junior clinicians, the main shortfall in stone knowledge seems to lie less in missing facts than in poor calibration and under-confidence, a problem that worsens as knowledge fades after training. Education built around cases and rich in feedback, sustained past graduation and working on factual grasp and on confidence calibration at the same time, might sharpen stone-related decisions and narrow the scope for avoidable error in how patients are counselled and cared for.

Background

Few urological complaints turn up as routinely in day-to-day practice as kidney stones. Across Western countries, the lifetime risk sits near 10%, and in the United States, it climbs past 12% in men and 6% in women.[1, 2] Diagnoses, too, have been mounting over recent decades.[1] What drives the rise is usually put down to low fluid intake and diet [3], set against a wider backdrop of obesity and metabolic syndrome. Recurrence is closer to the rule than the exception; somewhere between 30% and 40% of first-time formers produce a second stone within five years.[2] The cost is far from trivial, either. Stone disease has been rated the second most expensive urological condition, with its worldwide bill reaching roughly USD 5.3 billion in 2014.[4] Two compositions dominate what clinicians actually encounter: calcium oxalate (CaOx) and uric acid (UA). CaOx is the commonest form worldwide and within Saudi Arabia alike.[5] One local quirk stands out, though. UA stones make up about 10% of analyzed specimens in the Kingdom, a higher share than in many other regions, and this has been tied to the heavy local burden of diabetes and obesity.[5]

Within Saudi Arabia, the load is unusually heavy. Much of the blame falls on the hot, dry climate and the chronic dehydration that travels with it, alongside urbanization and a steadily more westernized diet.[6] One national survey put the prevalence among residents at roughly 9.1%, although the real figure is probably higher, since a fair number of cases never get picked up at the primary-care level.[7] Work in Riyadh flagged hypertension, a family history of stones, gout, and advancing age as the leading risk factors.[8] The disease also leans markedly male, with men accounting for close to 80% of the stones sent for analysis in one large Saudi series.[5]

Telling CaOx and UA stones apart matters in the clinic. They diverge in how they form and, more to the point, in how they respond to treatment. UA stones tend to crystallize in urine that stays acidic, frequently in the setting of metabolic syndrome or type 2 diabetes; CaOx stones, by contrast, stem largely from disordered handling of urinary calcium and oxalate.[1, 2] Alkalinizing the urine with oral citrate or bicarbonate will usually dissolve a UA stone and can spare the patient an operation.[4] CaOx stones behave differently. No drug reliably dissolves them, so management rests on spontaneous passage or surgery.[9] Imaging carries much of the diagnostic weight. Non-contrast computed tomography (NCCT) remains the reference standard for confirming stone disease.[10] When the scan is read, the density in Hounsfield units (HU) earns its keep, hinting at composition and at how the stone is likely to fare under extracorporeal shock wave lithotripsy (ESWL), even if the published thresholds drift from one study to the next.[11] Size is the other radiographic lever. It is usually taken as the longest two-dimensional diameter on axial slices, yet a three-dimensional read (whether volumetric or a three-dimensionally oriented multiplanar measurement of the longest axis) maps more faithfully onto the true stone burden and forecasts both treatment outcomes and later stone events more reliably.[12, 13] A clinician shaky on these fundamentals risks sending a patient for needless surgery, or letting a chance at non-invasive care slip past. Errors of that kind (needless procedures, imaging that was never indicated, dissolvable stones managed as though they were not) sit squarely within the domain of patient safety: they are avoidable, they follow from what the treating clinician knows rather than from any failure of

equipment or protocol, and they are therefore tractable through education. Whether that knowledge holds up across a career, once the structured teaching of medical school gives way to independent practice, is accordingly a patient-safety question as much as an educational one.

For all this prevalence, surprisingly little work has gauged what Saudi healthcare workers actually know about stones. A recent systematic review of nephrolithiasis management in the Middle East judged regional stone care to fall short, and it named gaps in workers' knowledge among the reasons.[14] Keeping primary-care doctors current on risk factors has itself been singled out as a lever against new and recurrent stones, which makes any knowledge deficit clinically meaningful.[6] What Saudi research does exist has mostly looked outward, toward patients and the public, for instance, measuring how well the general population recognizes stone risk factors (here, respondents with a medical background outscored the rest),[15] or probing the metabolic link between calcium oxalate stones and gout.[16] Studies that put the knowledge and attitudes of trainees and junior physicians to an objective test, however, are conspicuously thin on the ground.

In both primary care and the emergency department, it is usually the clinical-year student, the intern, or the general practitioner who meets the stone patient first. They order the first scan, they read it, and they make the call between conservative management and a urology referral. Should a gap open between what these clinicians believe they know and what they can actually demonstrate, the fallout can include misdiagnosis, imaging that was never needed, dissolvable stones left unrecognized, and referrals that arrive late. As far as we can tell, no published study has yet measured this perception-versus-reality gap among medical trainees and junior physicians in Saudi Arabia. Addressing that gap is what the present cross-sectional study set out to do: to assess and compare the factual and self-perceived knowledge of medical students, interns, and general practitioners in Riyadh around the differentiation and management of CaOx and UA stones, and to identify the factors that accompany adequate knowledge in this population.

Methods

Study design and sampling

This cross-sectional, descriptive survey was conducted in the Riyadh region of Saudi Arabia. Its purpose was to weigh the objective (factual) knowledge of clinical-year medical students, medical interns, and general practitioners against their self-perceived knowledge of how calcium oxalate and uric acid kidney stones differ, taking into account how well they grasped the pathophysiological and radiographic distinctions between the two, as well as the therapeutic ones.

Responses were gathered through a bilingual Arabic-and-English electronic survey hosted on Google Forms. Distribution ran through Prince Sultan Military Medical City (PSMMC) and affiliated institutions over two main channels, an institutional academic mailing list and a WhatsApp group, both well-worn routes for trainee-targeted survey work in the Saudi setting. Every Google Form submission counted as a single response. No email or IP addresses were captured, so duplicate entries could not be tracked by

technical means; repeat participation was nonetheless unlikely, given that each potential respondent was approached only once across these channels.

Sampling was by convenience. Of the 412 responses that first came in, incomplete entries were dropped, which left 407 participants for analysis, enough to support a prevalence estimate at a 95% confidence interval with a 5% margin of error.

This study is reported in accordance with the STRengthening the Reporting of OBservational studies in Epidemiology (STROBE) statement for cross-sectional studies; the completed checklist is provided in the **Supplementary File**.

Inclusion and exclusion criteria

Medical trainees and junior physicians practising or studying in Riyadh made up the target population. Eligibility was confined to three groups: clinical-year medical students, medical interns, and general practitioners. Preclinical students and specialists were left out, both to keep the data sound and to mark off a clear study population.

Questionnaire structure and scoring

Self-administered, the electronic form was built from the existing literature and checked through expert review by the research team. Electronic consent came first, then items on demographics, clinical background, and previous clinical exposure to kidney stones. A single item on years of experience was addressed to general practitioners but left open to every respondent, and was the only question that did not require an answer. Presenting each question in both languages was meant to keep a respondent's linguistic background from skewing the answers. The full English-language version of the instrument is provided in Supplementary File 1.

Thirteen five-point Likert items (running from strongly disagree to strongly agree) captured perceptions and beliefs. Eight of these formed a self-perceived knowledge and experience subscale and four an attitude and belief subscale, with a thirteenth item gauging how far respondents weighed differential diagnostic and therapeutic approaches.

Objective knowledge rested on four factual items: CT stone density, the best radiographic dimension to guide management, and the chemical dissolvability of calcium oxalate against uric acid stones. The raw score was then rescaled onto a 0–10 objective knowledge scale. Both factual and self-perceived scores were binned by modified Bloom's taxonomy thresholds, namely Poor (< 60%), Fair (60–79%), and Good (≥ 80%). A knowledge-gap analysis (self-perceived minus factual) was folded in as well, meant to surface mismatches between perceived expertise and measured clinical knowledge. Two urology consultant authors, FA and WK, vetted face and content validity and judged the questionnaire fit for distribution before data collection began. Internal consistency for the full thirteen-item self-perceived scale was good (Cronbach's $\alpha = 0.834$). One caveat should be flagged: the instrument went into full use without a separate pilot study beforehand.

Statistical analysis

Analyses were carried out in IBM SPSS Statistics (Version 26.0), treating a p-value < 0.05 as statistically significant. Categorical variables were summarised as frequencies and percentages, and Likert items as means with standard deviations. To compare the three cadres, the Kruskal-Wallis H test (with epsilon-squared as the effect size) was applied to the distributions of the Likert items and to the factual and self-perceived knowledge scores; wherever Kruskal-Wallis reached significance, Dunn-Bonferroni post-hoc pairwise comparisons followed.

For categorical comparisons across the three cadres, chi-square tests (with Cramér's V as the effect size) were used for Bloom's classifications along with the distribution of demographic and clinical-exposure characteristics. To gauge perception-versus-reality calibration, the link between the mean self-perceived subscale score and the total factual knowledge score was examined with both Pearson and non-parametric Spearman rank correlation coefficients.

Multivariable regression then served to pick out independent predictors of the knowledge outcomes. Candidate predictors were gender, age group, cadre, and prior exposure to kidney stone patients; years of clinical experience were set aside because of heavy missingness, the item having been optional and left unanswered by nearly half the sample. The continuous factual score and the knowledge gap (self-perceived minus factual) were each modelled by linear regression, the odds of reaching a Good factual classification by binary logistic regression, and, as a confirmatory check, the three Bloom's categories by ordinal logistic regression. Coefficients from all four models are reported together in the Results.

Ethical considerations

Approval for the study came from the Institutional Review Board (IRB) of Prince Sultan Military Medical City (Approval No. E-2959), and the work adhered closely to the ethical principles of the Declaration of Helsinki. Before any data were collected, mandatory electronic informed consent was obtained from every participant on the survey's opening page. At no stage were personal identifiers gathered, whether names, national or hospital identification numbers, email addresses, or IP addresses, and all responses were stored in an invite-protected account open only to named members of the research team. Every contributing author, moreover, had completed formal research bioethics training on the National Committee of Bioethics (NCBE) platform and holds an active NCBE certification at the King Abdulaziz City for Science and Technology (KACST). A large language model (LLM) was used solely to check grammar and English-language expression; it was not used to design the study, generate or analyze data, or interpret findings, and all content was reviewed and approved by the authors.

Results

1. Demographic Characteristics

The questionnaire was completed by 407 healthcare participants (Table 1). Men predominated (n = 314, 77.1%) over women (n = 93, 22.9%). Younger respondents weighted the age distribution: 79.6% (n = 324) fell in the 18–25 band, 18.4% (n = 75) in the 26–33 band, and just 2.0% (n = 8) in the 34–40 band. Medical students made up more than half of the cohort (n = 213, 52.3%), with interns (n = 122, 30.0%) and general practitioners (n = 72, 17.7%) behind them; no specialists or consultants were included. Among the 217 participants (53.3%) who answered the optional item on years of experience, responses clustered at the junior end: 76.0% with under 2 years, 18.0% with 2–4 years, 5.1% with 5–8 years, and 0.9% with 9–12 years. A shade over half of the sample had attended to at least one patient with kidney stones (n = 211, 51.8%; Table 1). Cadres diverged significantly on each demographic axis examined: gender ($\chi^2(2) = 23.24$, $p < .001$), age group ($\chi^2(4) = 144.41$, $p < .001$), and exposure to kidney stone patients ($\chi^2(2) = 35.19$, $p < .001$). Exposure was reported most often by general practitioners (75.0%), then interns (61.5%), then medical students (38.5%).

Table 1
Demographic characteristics of the study sample (N = 407).

Variable	**Category**	**n**	**%**
Gender	Male	314	77.1
	Female	93	22.9
Age Group	18–25	324	79.6
	26–33	75	18.4
	34–40	8	2.0
Cadre	Medical Student	213	52.3
	Intern	122	30.0
	General Practitioner	72	17.7
Years of Experience (n = 217 with reported experience)	< 2	165	76.0
	2–4	39	18.0
	5–8	11	5.1
	9–12	2	0.9
Attended Stone Patients	Yes	211	51.8
	No	196	48.2

2. Reliability Statistics

Internal consistency for the 13-item self-perceived knowledge scale was good; its Cronbach's alpha of 0.834 suggests that the items hung together and tapped a single underlying construct.

3. Likert Distributions and Bloom's Classification

Self-perception across the 13 Likert items came out mixed (Table 2, Panel A). Means ran from 2.69 ± 1.11 (Q1, experience in dealing with kidney stones) up to 3.88 ± 1.07 (Q5, the clinical importance of telling stone types apart). Agreement was strongest on items about clinical importance and basic risk factors; questions that probed the structural and physiological make-up of stone subtypes drew more neutral or dissenting answers, which hints at thinner self-confidence in the deeper pathophysiology.

Scored against Bloom's cut-offs (Good ≥ 80%, Fair 60–79%, Poor < 60%), the factual results split into 41.3% (n = 168) Good, 32.4% (n = 132) Fair, and 26.3% (n = 107) Poor. Self-perceived classification told a very different story: a mere 6.6% (n = 27) placed themselves in the Good band, 29.7% (n = 121) in Fair, and 63.6% (n = 259) in Poor (Table 2, Panel B).

That gap between measured and perceived standing, a factual mean of 6.70 ± 2.11 (67.0% correct) set against a self-perceived mean of 5.29 ± 1.92, points to systematic under-confidence (mean Knowledge Gap = − 1.41 ± 2.49): participants tended to sell their own factual knowledge short.

Table 2
Self-perceived knowledge item distributions and Bloom's modified cut-off classification.

Panel A. *Mean Likert scores for self-perception items (1 = Strongly Disagree, 5 = Strongly Agree).*

Item	Mean	SD
Q1. Experienced in dealing with kidney stones	2.69	1.11
Q2. Stone formation more in men than women	3.36	1.04
Q3. Obesity and inactivity increase risk	3.86	0.99
Q4. Understand pathophysiological basis	3.25	1.04
Q5. Important to differentiate stone types clinically	3.88	1.07
Q6. Understand structural differences (CaOx vs UA)	3.35	1.08
Q7. Understand physiological differences	3.15	1.13
Q8. Calcium and uric acid increase formation	3.82	0.98
Q9. Calcium supplements increase risk	3.61	1.00
Q10. Understand mixed CaOx/UA stones	3.07	1.05
Q11. Apply pathophysiological knowledge in clerkship	3.19	0.97
Q12. Labs/images help determine stone type	3.87	1.07
Q13. Consider differential diagnostic approaches	3.46	0.91

Panel B. *Bloom's classification of factual versus self-perceived knowledge.*

Category	Factual Knowledge n (%)	Self-Perceived Knowledge n (%)
Poor (< 60%)	107 (26.3)	259 (63.6)
Fair (60–79%)	132 (32.4)	121 (29.7)
Good (≥ 80%)	168 (41.3)	27 (6.6)

4. Cadre Comparisons (Likert + Factual, with Effect Sizes)

Factual scores differed significantly across the three cadres (Kruskal-Wallis H = 7.995, df = 2, p = .018). The top mean went to medical students (6.95 ± 1.93), with interns (6.70 ± 2.13) and then general practitioners (5.99 ± 2.41) trailing (Table 3, Panel A). The finding runs against intuition: for all their extra clinical exposure, general practitioners landed lowest on the factual items.

On Dunn-Bonferroni post-hoc testing, medical students outscored general practitioners by a significant margin (adjusted p = .014). Neither the student-versus-intern nor the intern-versus-GP comparison reached significance.

Self-perception, too, varied significantly by cadre (Kruskal-Wallis H = 11.489, df = 2, p = .003). Interns rated themselves highest (5.72 ± 1.92), ahead of GPs (5.32 ± 2.09) and students (5.03 ± 1.82). The narrowest under-confidence gap, though, sat with general practitioners (− 0.67), then interns (− 0.98), the two groups that edged closest to overconfidence. Medical students went the other way. Top factual scores paired with the lowest self-perception, which produced the largest negative Knowledge Gap (− 1.92) and the sharpest underconfidence in the sample.

Five Likert items split significantly across cadres at the item level (Table 3, Panel B): Q1 (experience: H = 23.06, p < .001), Q6 (structural differences: H = 6.68, p = .036), Q7 (physiological differences: H = 6.38, p = .041), Q11 (utilization in clerkship: H = 9.24, p = .010), and Q13 (differential diagnostic approaches: H = 6.77, p = .034). The remaining items showed no cadre-related difference of note.

The spread of Bloom's categories shifted significantly with cadre for both factual knowledge ($\chi^2(4) = 11.38$, p = .023; Cramer's V = 0.118, small effect) and self-perceived knowledge ($\chi^2(4) = 11.68$, p = .020; Cramer's V = 0.120, small effect). A Poor factual rating was most common among GPs (41.7%) and rarest among students (21.6%).

Table 3

Knowledge comparisons by professional cadre.

Panel A. *Factual and self-perceived knowledge by cadre (mean ± SD; non-parametric tests).*

Variable	Student	Intern	GP	Test	p	Effect (ε^2)
Factual Knowledge	6.95 ± 1.93	6.70 ± 2.13	5.99 ± 2.41	H = 7.995	.018	0.020
Self-Perceived	5.03 ± 1.82	5.72 ± 1.92	5.32 ± 2.09	H = 11.489	.003	0.028
Knowledge Gap	−1.92 ± 2.55	−0.98 ± 2.49	−0.67 ± 2.06	N/A	N/A	N/A

Note. Post-hoc Dunn-Bonferroni analysis: medical students scored significantly higher than general practitioners (adjusted p = .014); other pairwise comparisons were non-significant.

Panel B. *Significant Likert items by cadre (Kruskal-Wallis).*

Item	H	df	p-value
Q1. Experience with stones	23.06	2	< .001
Q6. Structural differences	6.68	2	.036
Q7. Physiological differences	6.38	2	.041
Q11. Use in clerkship	9.24	2	.010
Q13. Differential approaches	6.77	2	.034

Note. H = Kruskal-Wallis test statistic comparing the three cadres (medical students, interns, and general practitioners); df = 2. Only items reaching statistical significance are listed.

5. Factual Knowledge % Correct

Expressed as percentages, mean factual accuracy fell as clinical seniority rose, an inverse gradient across the three cadres. Students reached 69.5% correct (95% CI: 66.9–72.1), interns 67.0% (95% CI: 63.2–70.8), and general practitioners 59.9% (95% CI: 54.2–65.5).

Broken down by item (Q14-Q17), factual accuracy swung widely from one domain to the next (Fig. 1). The strongest showings came on the role of CT density in management (88.4% correct) and on the dissolvability of uric acid stones (82.8% correct). Knowledge thinned out elsewhere: only 42.1% recognized that calcium oxalate stones will not dissolve pharmacologically, and 53.5% picked the optimal radiographic dimension for management.

6. Self-Perceived vs. Factual Knowledge

Self-perceived and factual knowledge were related through both Pearson and Spearman correlations. The link proved positive, significant, but weak (Pearson $r = 0.237$, $p < .001$; Spearman $\rho = 0.224$, $p < .001$), so self-assessment seems to capture only a sliver of the real variance in knowledge ($R^2 \approx 5.6\%$). The Bloom's discrepancy (Table 2, Panel B) drives the point home, with 41.3% of participants showing Good factual knowledge but only 6.6% claiming it for themselves, a pattern that reads as systematic under-confidence rather than accurate self-appraisal. One might see a Dunning-Kruger-like signature here: the weakest performers tended toward positive gaps (overestimating), while the strongest tended to undersell themselves. Across the sample, the mean Knowledge Gap came to -1.41 ± 2.49, confirming under-confidence overall, even though a handful of low scorers ran the other way. Stratified by cadre, the largest negative gap belonged to medical students (-1.92 ± 2.55), a marked under-confidence; GPs sat at the other end with the smallest gap (-0.67 ± 2.06), tracking their actual performance more closely, if still on the negative side.

7. Predictive Models of Knowledge Outcomes

To pin down independent predictors of the knowledge outcomes, four multivariable regression models were fitted (N = 407). Each took gender, age group, cadre, and exposure to kidney stone patients as predictors. Years of experience were dropped, given 47% missingness, since the item was optional. Table 4 collects the coefficients from all four models.

Table 4

Multivariable predictors of knowledge outcomes.

Predictor	Model 1 Factual score β [95% CI], p	Model 2 Good vs Not-Good OR [95% CI], p	Model 3 Bloom's category OR [95% CI], p	Model 4 Knowledge Gap β [95% CI], p
Female (ref: Male)	−0.11 [− 0.60, 0.38], 0.648	0.85 [0.52, 1.40], 0.532	0.92 [0.59, 1.45], 0.731	0.21 [− 0.38, 0.80], 0.477
Age 26–33 (ref: 18–25)	−0.53 [− 1.14, 0.08], 0.091	0.68 [0.36, 1.27], 0.230	0.71 [0.40, 1.26], 0.243	0.27 [− 0.46, 1.01], 0.465
Age 34–40 (ref: 18–25)	**−3.63 [− 5.16, − 2.10], < .001**	Not estimable□	**0.06 [0.01, 0.29], 0.001**	1.51 [− 0.33, 3.34], 0.107
Intern (ref: Student)	−0.27 [− 0.76, 0.22], 0.287	0.99 [0.60, 1.61], 0.957	0.93 [0.59, 1.47], 0.765	**0.89 [0.30, 1.48], 0.003**
GP (ref: Student)	−0.50 [− 1.18, 0.19], 0.157	0.89 [0.44, 1.79], 0.743	0.72 [0.38, 1.39], 0.327	**0.93 [0.11, 1.76], 0.027**
Attends patients (ref: No)	0.38 [− 0.04, 0.80], .075	1.14 [0.75, 1.72], 0.549	1.23 [0.84, 1.82], 0.289	0.19 [− 0.31, 0.69], 0.453

Note. Models 1 and 4 report β (linear regression); Models 2 and 3 report OR (binary and ordinal logistic regression). All models were adjusted for gender, age group, cadre, and patient exposure, each coefficient being relative to the reference category shown in parentheses. Years of experience were excluded (47% missing). Bold indicates p < .05. □ Omitted due to complete separation.

7.1 Model 1, Linear Regression: Factual Knowledge Score (0–10)

Linear regression drove Model 1, predicting the continuous factual score. Age dominated. Those in the 34–40 band scored 3.63 points below the 18–25 group (β = −3.63, 95% CI: −5.16 to − 2.10, p < .001), a wide and clinically meaningful gap. A borderline positive signal came from exposure to kidney stone patients (β = 0.38, 95% CI: −0.04 to 0.80, p = .075). After adjustment, gender, the 26–33 band, and cadre carried no predictive weight.

7.2 Model 2, Binary Logistic Regression: Good (≥ 80%) vs. Not-Good Knowledge

Model 2 turned to binary logistic regression for the odds of Good factual knowledge (≥ 80% correct) against Not Good. The 34–40 age category had to be dropped here because of complete separation: every one of its 8 participants fell into Not-Good, so maximum likelihood would not converge at that level. As a sensitivity analysis, Firth's penalized regression is the sensible next step. None of the surviving predictors held up after adjustment. Odds ratios for cadre, gender, and exposure to kidney

stone patients all hover near 1.0 on wide confidence intervals, a sign that demographics alone do little to explain the binary Good/Not-Good split.

7.3 Model 3, Ordinal Logistic Regression: Bloom's Three-Level Classification

For Model 3, ordinal (proportional-odds) logistic regression predicted the three-level Bloom's classification (Poor / Fair / Good). The standout result was a strong, significant inverse tie to the oldest band: at ages 34–40, the odds of sitting in a higher Bloom's category were just 6% of those at 18–25 (OR = 0.06, 95% CI: 0.01–0.29, p = .001). That echoes the inverse seniority-knowledge gradient seen in the unadjusted cadre comparisons (Section 4). Nothing else reached significance, which suggests it is age, not cadre as such, that pushes the ordinal shift toward lower Bloom's categories.

7.4 Model 4, Linear Regression: Knowledge Gap (Self-Perceived − Factual)

Model 4 again used linear regression, this time on the standardized Knowledge Gap (self-perceived minus factual), where negative values flag under-confidence and positive values over-confidence. With age, gender, and patient exposure held constant, professional cadre emerged as a significant predictor of calibration: interns (β = 0.89, 95% CI: 0.30–1.48, p = .003) and general practitioners (β = 0.93, 95% CI: 0.11–1.76, p = .027) were both far less under-confident than medical students. So the calibration of self-perception appears to sharpen as clinical role advances, even where factual knowledge does not, which fits a Dunning-Kruger-type pattern at the early-training end of the spectrum. Age, gender, and patient exposure each failed to predict the gap on their own.

Read together, the four models tell a fairly coherent story. Age, specifically the 34–40 band, pulls factual knowledge down in both its continuous and ordinal forms, whereas cadre governs the calibration of self-perception, with interns and GPs better calibrated than students despite no edge in factual scores. In none of the models did gender or patient exposure stand on its own as a predictor.

Discussion

What this study brings to light is a clinically important gap between what medical trainees and junior physicians know about urinary stone management and what they believe they know, with under-confidence, not over-confidence, the recurring theme. Most writing on medical self-assessment focuses on over-confidence, yet under-confidence carries comparable risk at the front line, all the more so in a high-prevalence condition like nephrolithiasis, where junior clinicians so often make the first assessment and referral call. The gap showed most plainly in one contrast: far more participants demonstrated "good" factual knowledge than rated their own competence as good (41.3% vs 6.6%), and across the board, people scored themselves below what they actually achieved. In day-to-day terms, that hesitation can spill into delayed escalation, tests that were never warranted, or the underuse of conservative routes such as urine alkalinization for uric acid stones.

This split between confidence and competence sits comfortably with a large body of work showing self-assessment to be a shaky guide to actual ability. Here, the correlation between self-perceived and factual knowledge was weak, though statistically significant, and explained only a small part of the variance in objective performance. Since the self-perception scale held together well (Cronbach's α = 0.834 across 13 items), the underconfidence is probably not just a measurement artifact. Davis et al., reviewing the field, reached a similar verdict: most comparisons between physician self-assessment and outside measures of competence showed little, no, or even an inverse relationship, and the least-skilled, most-confident physicians made the least accurate judges of themselves [17]. On balance, self-perception looks to be moulded more by educational climate, by feedback, and by one's comfort with uncertainty than by competence as such.

Notably, the calibration seen here was not a textbook Dunning-Kruger effect [18, 19]. The lowest performers leaned toward overestimation, yet underconfidence ruled overall and bit hardest among the stronger performers. The shape recalls the impostor-like calibration so often reported in medical learners, who talk down their competence as tasks grow more complex or uncertain [20]. It also chimes with the idea of "conscious incompetence," whereby more advanced learners grow ever more alert to clinical nuance and, in turn, less sure of what they know [21]. Tasks that look simple on paper, such as categorizing stone composition or reading CT density, or settling on a management pathway, turn out to be a good deal harder in practice, and that distance may feed the reflective uncertainty.

The cadre comparisons brought this into sharper focus. Medical students posted numerically higher factual knowledge than both interns and general practitioners, though the difference was significant only against general practitioners (adjusted $p = .014$); even so, they rated themselves lowest, which generated the largest negative knowledge gap. General practitioners scored lowest on the factual items in spite of the most patient exposure, while interns rated themselves highest. Modelling age, gender, cadre, and exposure together changed the reading, though: the factual shortfall traced back to age, not cadre. Participants aged 34–40 scored well below the rest ($\beta = -3.63$, 95% CI − 5.16 to − 2.10, $p < .001$) and faced much lower odds of a higher Bloom's category (OR = 0.06, 95% CI 0.01–0.29, $p = .001$), though with only eight people in the band, this needs caution. The finding fits the well-documented decay of theoretical knowledge after medical school, once retrieval practice and steady reinforcement taper off [22]. Calibration, by contrast, was the part cadre predicted. Interns ($\beta = 0.89$, $p = .003$) and general practitioners ($\beta = 0.93$, $p = .027$) were both significantly less under-confident than students, so self-assessment accuracy seems to climb with clinical role, even when factual knowledge does not; the students' under-confidence despite strong performance echoes reports that the young physicians rated highest by observers tend to rate their own self-efficacy lowest [23].

At the item level, the unevenness carried clinical meaning. Participants handled the role of CT stone density in management and the dissolvability of uric acid stones well, but stumbled on the non-dissolvability of calcium oxalate stones and on the best radiographic dimension for planning (Fig. 1). Comfort with Hounsfield-unit density probably comes from its everyday use in reading NCCT and its worth in forecasting stone fragility and ESWL success [24]. The shakier grip on calcium oxalate non-

dissolvability is the more worrying of the two, since the misconception can breed false hope of pharmacological dissolution and stall procedural therapy when spontaneous passage is improbable [25]. Locally, this matters all the more because calcium oxalate is the commonest subtype in the Saudi series, roughly 74% of pure-oxalate stones in the Eastern Province and around 85% in a larger regional cohort [5, 26], so a thin grasp of its management bears straight on the stone clinicians meet most. The flip side, strong recognition that uric acid stones dissolve, is locally apt, given that uric acid stones run proportionally higher in Saudi Arabia than in many regions, a pattern tied to the heavy burden of diabetes and obesity [5]. One detail stands out: the lowest self-rated items, hands-on experience and the deeper pathophysiology of stone subtypes (grasping mixed calcium-oxalate/uric-acid stones, mean 3.07; physiological differences between subtypes, 3.15), landed on the same ground as the weakest factual domain, which says that low confidence and weak knowledge meet over subtype pathophysiology rather than over clinical fundamentals.

Pulled together, the results imply that the core problem in nephrolithiasis training is less a raw deficit of facts than a failure to align competence with confidence, made worse by the knowledge that slips away once the exam-driven phase of training ends. Teaching, then, needs to reach past content delivery. Structured feedback, case-based work, hands-on radiological interpretation, and explicit calibration exercises that help learners size up their own competence all have a place. Continuing medical education aimed squarely at interns and general practitioners might also blunt the post-graduation drop in retention and build knowledge and self-awareness together in stone management.

A few limitations deserve mention. Being cross-sectional, the design cannot establish cause between educational exposure and knowledge outcomes, and leaning partly on self-report invites response bias. With the sample tilted toward younger participants and medical students, the findings may not carry over to more senior physicians or specialists, and the factual test rested on only a handful of items that may miss the wider sweep of competence. One more point: although several between-cadre comparisons were significant, the effect sizes stayed small ($\varepsilon^2 \approx 0.02–0.03$ for factual and self-perceived scores; Cramér's $V \approx 0.12$ for the Bloom's-category distributions), so cadre explained only a modest slice of the variance, and the differences ought to be read in that light.

Conclusions

In sum, the study points to a systematic mismatch between perceived and actual knowledge of urinary stone management among medical trainees and junior physicians, one driven more by miscalibration than by any real shortage of facts. Building up factual understanding alongside confidence calibration, through continuous education that is rich in feedback and grounded in cases, may sharpen clinical decisions and improve both how patients are counselled and how stone disease is managed overall.

Abbreviations

CaOx

Calcium oxalate

CI

Confidence interval

CT

Computed tomography

ESWL

Extracorporeal shock wave lithotripsy

GP

General practitioner

HU

Hounsfield unit

IRB

Institutional Review Board

KACST

King Abdulaziz City for Science and Technology

LLM

Large language model

NCBE

National Committee of Bioethics

NCCT

Non-contrast computed tomography

OR

Odds ratio

PSMMC

Prince Sultan Military Medical City

SD

Standard deviation

SPSS

Statistical Package for the Social Sciences

UA

Uric acid

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of Prince Sultan Military Medical City, Riyadh, Saudi Arabia (Approval No. E-2959), and was conducted in accordance with the Declaration of Helsinki. Electronic informed consent was obtained from all participants before data collection.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

RA: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Project administration, Writing – original draft. RMN: Conceptualization, Methodology, Resources, Supervision, Project administration, Writing – review & editing. AMA: Investigation, Data curation, Writing – review & editing. SA: Investigation, Resources, Writing – review & editing. KAlm: Investigation, Validation, Writing – review & editing. KAlk: Investigation, Data curation, Writing – review & editing. FA: Investigation, Writing – review & editing. WK: Validation, Writing – review & editing. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

References

1. Cloutier J, Villa L, Traxer O, Daudon M. Kidney stone analysis: Give me your stone, I will tell you who you are! World J Urol. 2015;33(2):157–69. https://doi.org/10.1007/s00345-014-1444-9.
2. Curhan GC. Epidemiology of stone disease. Urol Clin North Am. 2007;34(3):287–93. https://doi.org/10.1016/j.ucl.2007.04.003.
3. Friedlander JI, Antonelli JA, Pearle MS. Diet: from food to stone. World J Urol. 2015;33(2):179–85. https://doi.org/10.1007/s00345-014-1344-z.

4. Akram M, Jahrreiss V, Skolarikos A, Geraghty R, Tzelves L, Emilliani E, et al. Urological guidelines for kidney stones: overview and comprehensive update. J Clin Med. 2024;13(4):1114. https://doi.org/10.3390/jcm13041114.
5. Amir A, Matlaga BR, Ziemba JB, Sheikh S. Kidney stone composition in the Kingdom of Saudi Arabia. Clin Nephrol. 2018;89(5):345–8. https://doi.org/10.5414/cn109313.
6. Alblowi S, Safdar O, Aboulola N, Alharazy D, Najem N. Renal stone prevalence and risk factors in Jeddah and Riyadh. J Family Med Prim Care. 2022;11(6):2839–45. https://doi.org/10.4103/jfmpc.jfmpc_511_21.
7. Safdar OY, Alzahrani WA, Kurdi MA, Ghanim AA, Nagadi SA, Alghamdi SJ, et al. The prevalence of renal stones among local residents in Saudi Arabia. J Family Med Prim Care. 2021;10(2):974–7. https://doi.org/10.4103/jfmpc.jfmpc_262_20.
8. Alshubaili AM, Alotaibi AF, Alsaleh KA, Almogarri AI, Alanizi AA, Alsaif SS, et al. The prevalence of nephrolithiasis and associated risk factors among the population of the Riyadh Province, Saudi Arabia. Cureus. 2024;16(3):e55870. https://doi.org/10.7759/cureus.55870.
9. Malieckal DA, Ganesan C, Arellano Mendez D, Pao AC. Breaking the cycle of recurrent calcium stone disease. Adv Kidney Dis Health. 2023;30(2):164–76. https://doi.org/10.1053/j.akdh.2022.12.004.
10. Haghighi R, Rezaei J, Khaleghi F, Hamidi Madani M, Soleimantabar H, Norouzi M, et al. Preferred radiological modalities in the diagnosis and management of kidney and urinary tract stones; a systematic review. J Ren Inj Prev. 2025;14(2):e38396. https://doi.org/10.34172/jrip.2025.38396.
11. Garg M, Johnson H, Lee SM, Rai BP, Somani B, Philip J. Role of Hounsfield unit in predicting outcomes of shock wave lithotripsy for renal calculi: outcomes of a systematic review. Curr Urol Rep. 2023;24:173–85. https://doi.org/10.1007/s11934-023-01145-w.
12. Planz VB, Posielski NM, Lubner MG, Li K, Chen GH, Nakada SY, et al. Ultra-low-dose limited renal CT for volumetric stone surveillance: advantages over standard unenhanced CT. Abdom Radiol (NY). 2019;44(1):227–33. https://doi.org/10.1007/s00261-018-1719-5.
13. Reimer RP, Klein K, Rinneburger M, Zopfs D, Lennartz S, Salem J, et al. Manual kidney stone size measurements in computed tomography are most accurate using multiplanar image reformatations and bone window settings. Sci Rep. 2021;11:16437. https://doi.org/10.1038/s41598-021-95962-z.
14. Ahmed EM. Management of nephrolithiasis in the Middle East over a recent decade: a systematic review. Urol Ann. 2024;16(1):36–42. https://doi.org/10.4103/ua.ua_111_23.
15. AlOtipi AM, AlQurashi MM, AlTalhi AA, AlOtaibi AOA, AlThomali IHM, AlKhudaydi MS. Public awareness toward kidney stones risk factors in Saudi Arabia; a cross-sectional observational study. Int J Med Dev Ctries. 2020;4(1):101. https://doi.org/10.24911/IJMDC.51-1571682346.
16. Fazil Marickar YM. Calcium oxalate stone and gout. Urol Res. 2009;37(6):345–7. https://doi.org/10.1007/s00240-009-0218-0.
17. Davis DA, Mazmanian PE, Fordis M, Harrison RV, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. JAMA. 2006;296(9):1094–102. https://doi.org/10.1001/jama.296.9.1094.

18. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999;77(6):1121–34. https://doi.org/10.1037//0022-3514.77.6.1121.
19. Rahmani M. Medical trainees and the Dunning-Kruger effect: when they don't know what they don't know. J Grad Med Educ. 2020;12(5):532–4. https://doi.org/10.4300/JGME-D-20-00134.1.
20. Rosenthal S, Schlussel Y, Yaden MB, DeSantis J, Trayes K, Pohl C, et al. Persistent impostor phenomenon is associated with distress in medical students. Fam Med. 2021;53(2):118–22. https://doi.org/10.22454/fammed.2021.799997.
21. Eva KW, Regehr G. I'll never play professional football and other fallacies of self-assessment. J Contin Educ Health Prof. 2008;28(1):14–9. https://doi.org/10.1002/chp.150.
22. Maximous SI, Acho M, Ahari J, Hawkins K, Jamieson DB, Luks AM, et al. Decay in physiologic knowledge since medical school among critical care fellows (DIP study). ATS Sch. 2025;6(2):191–201. https://doi.org/10.34197/ats-scholar.2024-0036OC.
23. Gude T, Finset A, Anvik T, Bærheim A, Fasmer OB, Grimstad H, et al. Do medical students and young physicians assess reliably their self-efficacy regarding communication skills? BMC Med Educ. 2017;17:107. https://doi.org/10.1186/s12909-017-0943-y.
24. Kambadakone AR, Eisner BH, Catalano OA, Sahani DV. New and evolving concepts in the imaging and management of urolithiasis: urologists' perspective. Radiographics. 2010;30(3):603–23. https://doi.org/10.1148/rg.303095146.
25. Skolarikos A, Davis NF, Gambaro G, Geraghty R, Jung H, Kamphuis GM et al. EAU guidelines on urolithiasis. European Association of Urology; 2026. Available from: http://uroweb.org/guideline/urolithiasis/. Accessed 21 April 2026.
26. Alkhunaizi AM. Urinary stones in Eastern Saudi Arabia. Urol Ann. 2016;8(1):6–9. https://doi.org/10.4103/0974-7796.164841.

Figures

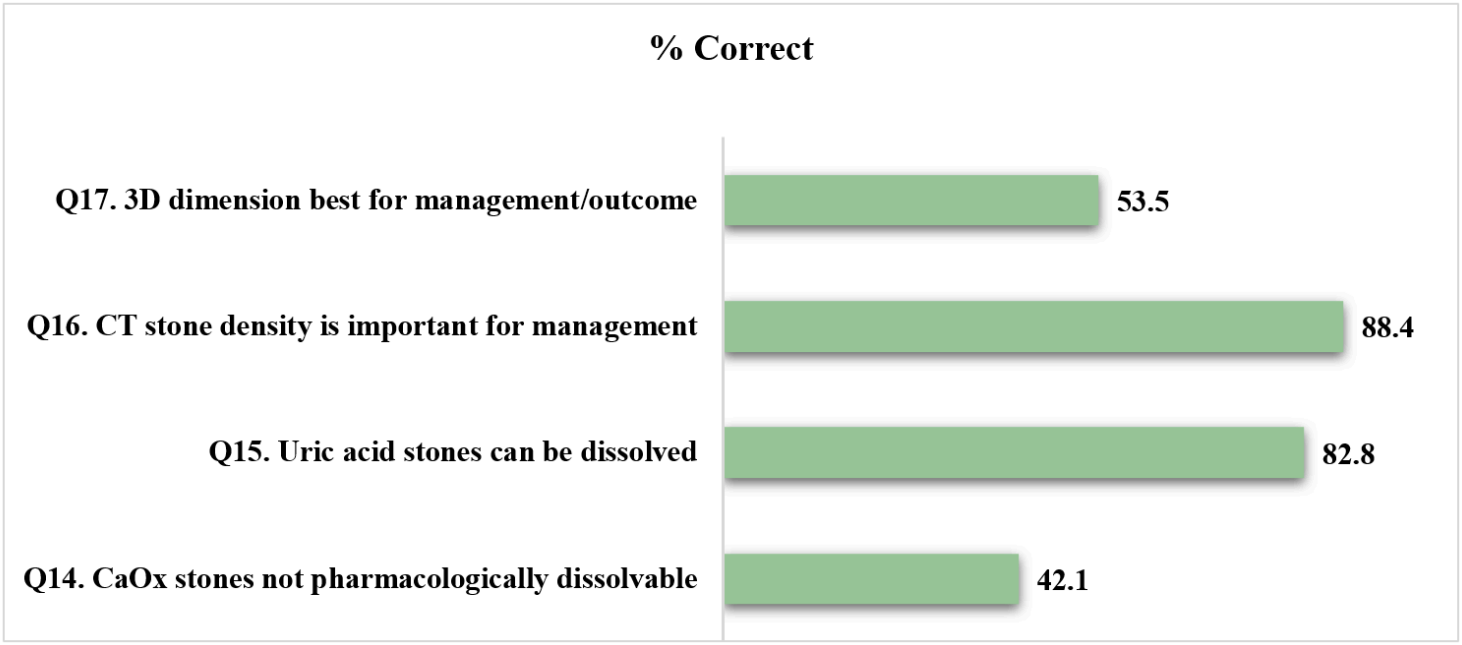


Figure 1

Item-level factual knowledge accuracy percentages for Q14-Q17.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFile1Questionnaire.docx