Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Minimum Procedural Volume Thresholds for Surgical Privileging: A Mixed-Methods Validation and Risk Management Framework in a Multi-Specialty Healthcare Network

Nantana Suppapitnarm [1], Suthorn Bavonratanavech [2,3], Wanicha Panyakhamlerd [4], Pongtorn Kietdumrongwong [5]

[1]Medical Affairs Division, Bangkok Dusit Medical Services PCL, Bangkok, Thailand; [2]Trauma and Orthopedics Center, Bangkok Dusit Medical Services PCL, Bangkok, Thailand; [3]Department of Orthopedics, Bangkok Hospital Headquarters, Bangkok, Thailand; [4]Women Center, Bangkok Hospital Headquarters, Bangkok, Thailand; [5]Health Research Center, Bangkok Dusit Medical Services PCL, Bangkok, Thailand

Correspondence: Nantana Suppapitnarm, Medical Affairs Division, Bangkok Dusit Medical Services PCL, 2 Soi Soonvijai 7, New Petchburi Road, Bang Kapi Subdistrict, Huai Khwang District, Bangkok, 10310, Thailand, Email nantana.su@bdms.co.th

Purpose: To establish evidence-based annual procedural volume thresholds and surgical pairing standards for Orthopedics and Obstetrics-Gynecology (OB-GYN), and to validate these benchmarks against clinical incident rates as a proactive risk management tool.

Methods: This study utilized a sequential mixed-methods design. In Phase 1, a Delphi technique was employed with 36 senior experts from the Bangkok Dusit Medical Services (BDMS) network to reach a consensus on minimum annual volumes and co-surgeon requirements. In Phase 2, a retrospective analysis of clinical incidents (Levels 4–5 and Sentinel Events) from 2022 to 2024 was performed to validate these thresholds. Statistical analysis included Mann–Whitney U and Kruskal–Wallis tests to evaluate the association between sustained compliance with these standards and clinical incident rates.

Results: Expert consensus established a recommended safety threshold of 10 cases per year for high-volume procedures (eg, cesarean delivery, anterior cruciate ligament [ACL] reconstruction), with specific thresholds of 6 cases for proximal humerus fracture and 5 cases for shoulder rotator cuff repair. Quantitative validation demonstrated that surgeons meeting these thresholds had significantly lower incident rates in cesarean delivery ($p < 0.001$), proximal humerus fractures ($p = 0.027$), and shoulder rotator cuff repair ($p < 0.001$). Furthermore, sustained multi-year compliance over three consecutive years was strongly correlated with lower incident rates ($p < 0.001$). For high-complexity, low-volume procedures, the strategy shifted from numerical frequency to system-based redundancy, requiring a mandatory co-surgeon (Median Score = 5.0 on a 5-point scale).

Conclusion: Maintaining a minimum annual procedural volume serves as a core proxy for surgical competency, directly reinforcing the knowledge, skill, and attitude required for patient safety. These established thresholds and team-based safety models provide a data-driven framework for institutional clinical privileging, offering hospital administrators an objective policy tool for proactive risk management.

Keywords: physician competency, clinical privileging, patient safety, delphi technique, orthopedic surgery, obstetrics and gynecology

Introduction

A review of clinical incidents within the Bangkok Dusit Medical Services (BDMS) network between 2022 and 2024 revealed severe incident rates of 2.65%, 2.48%, and 2.24%, respectively. Within the BDMS Enterprise Risk Management system (BDMS ERMs), these clinical incidents are classified by severity. Levels 4 and 5 denote serious adverse events leading to prolonged hospitalization or severe reputational and legal harm, while Sentinel Events (SE) represent catastrophic, unexpected outcomes—such as death, permanent harm, or severe temporary harm—requiring urgent reporting within 24 hours.

Crucially, a retrospective analysis identified physician competency—specifically, insufficient procedural volume and a lack of updated clinical knowledge—as the primary driver behind these severe events, accounting for 69.88%, 44.44%, and 24.14% of cases annually.

https://doi.org/10.2147/RMHP.S608778
Received: 16 March 2026
Accepted: 3 June 2026
Published: 17 June 2026

Historically, these competency-driven incidents have been most prevalent within Orthopedics, Obstetrics-Gynecology (OB-GYN), and the broader surgical field, which together constitute the top three clinical incident categories in our network. While the surgical field ranks high in total incidents, it is highly fragmented across multiple independent specialty boards (eg, urology, thoracic surgery, etc).

In contrast, Orthopedics and OB-GYN represent unified, high-volume, single-board surgical domains. This unified structure makes them ideal touchpoints for clinical risk management and provides a methodologically sound basis for validating standardized volume thresholds and institutional competency frameworks.

Physician competency integrates the knowledge, skills, and experience essential for professional medical practice. The pedagogical framework "Learn, See, Practice, Prove, Do, and Maintain," proposed by Sawyer et al,[1] emphasizes that procedural skills inevitably decline a phenomenon known as skill decay if not regularly practiced or formally re-assessed Consequently, establishing robust mechanisms for skill maintenance is paramount for ensuring patient safety.

While countries such as Australia,[2] the UK,[3] Canada,[4] and the US[5–8] have established competency benchmarks for residents and medical students,[9–12] a significant gap remains in standardized mandates for "skill maintenance" among practicing specialists. Similarly, while curricula in Germany,[13] Switzerland,[14] and the Netherlands[12] define specific procedural mastery, and the European Board and College of Obstetrics and Gynecology (EBCOG) PACT curriculum[15] provides standardized academic training across Europe, Thailand particularly its private healthcare sector currently lacks standardized minimum volume requirements for professional competency maintenance.

Problem Statement

The core problem lies in the absence of objective, evidence-based criteria for ongoing surgical privileging. In professional practice, clinicians rarely acknowledge a decline in their own technical competency, making subjective self-assessment an unreliable safeguard. Without concrete, data-driven thresholds linking clinical volume to patient outcomes, healthcare institutions face immense challenges in enforcing competency guidelines or restricting privileges without appearing arbitrary. To bridge this gap, objective and empirical evidence is urgently needed to reduce volume-related adverse events and systematically maintain physician competence.

Research Objectives

1. Establish consensus-based, recommended annual procedural volume thresholds and co-surgeon requirements for Orthopedics and OB-GYN using a multi-specialty expert Delphi panel.
2. Quantitatively validate these thresholds against retrospective clinical incident data (Levels 4–5 and Sentinel Events) within a multi-hospital private network.
3. Provide a validated, proactive risk management framework to guide institutional clinical privileging policies and enhance patient safety.

Materials and Methods

Study Design

This study utilized a sequential mixed-methods approach to establish and validate surgical competency standards. The research comprised two phases: (1) a qualitative consensus-building phase using an iterative Delphi technique, and (2) a quantitative validation phase through a retrospective clinical incident analysis. The study was conducted across the Bangkok Dusit Medical Services (BDMS) network and received ethical approval from the Institutional Review Board (IRB No. 2024–05-20).

Phase 1: Delphi Technique and Expert Panel

Expert Selection and Sample Size Justification

A purposive sampling strategy was employed to recruit an expert panel from Orthopedics and Obstetrics-Gynecology (OB-GYN). Inclusion criteria were: (1) board-certified specialists practicing within the BDMS network; (2) a minimum of six years of post-residency experience (Specialist or Senior Specialist rank); and (3) leadership or academic standing, including committee representatives from the Royal Thai College of Obstetricians and Gynecologists (RTCOG) or The

Royal College of Orthopedic Surgeons of Thailand (RCOST). To ensure statistical reliability, a target sample size of at least 17 experts per specialty group was established based on Macmillan (1971),[16] who demonstrated that once an expert panel reaches 17 participants, the marginal reduction in group error becomes minimal, ensuring consensus stability. To account for potential attrition and proactively cushion against dropouts, additional experts were initially invited. Ultimately, out of 37 invited experts, 36 completed all three Delphi rounds (17 Orthopedists and 19 Obstetricians), resulting in an excellent 97.3% retention rate.

Delphi Process and Consensus Operationalization

The Delphi process was executed over three sequential rounds, structured around four core methodological pillars to minimize selection and conformity biases:

1. **Anonymity**: Responses were collected via separate, non-identifiable electronic questionnaires. Panelists evaluated proposals independently, mitigating the risk of professional hierarchy or dominant personalities biasing the consensus.
2. **Iteration**: The same fundamental queries were reassessed across three rounds, allowing experts to refine their judgments without loss of professional face.
3. **Controlled Feedback**: Between rounds, the research team synthesized the inputs, providing participants with an organized summary of comments and rationales from both conforming and non-conforming perspectives before the next round.
4. **Statistical Group Response**: Feedback in Rounds 2 and 3 was explicitly presented using descriptive statistics, including group medians and dispersion metrics.

In Round 1, experts received a semi-structured questionnaire divided into three sections: (1) demographic profiles, (2) open-ended inquiries regarding clinical factors influencing long-term competency and the mandatory annual procedural volumes required to sustain technical proficiency, and (3) a prompt to propose additional high-risk procedures for a comprehensive registry. To ground volume estimations in empirical data, experts were provided with objective historical benchmarks from the network's 3-year retrospective data (2022–2024), including baseline procedural averages and relevant global literature.

In Rounds 2 and 3, panelists rated these synthesized items on a 5-point Likert scale. Final dispersion and consensus were measured using the Interquartile Range ($IQR \leq 1.5$), while agreement stability across rounds was monitored by tracking changes in median values and percent agreement to ensure inter-rater reliability.

Phase 2: Clinical Validation (Association Analysis)

Data Source and Outcomes

Procedural logs and clinical incident reports (2022–2024) were extracted from the BDMS centralized database. Clinical incidents were strictly categorized based on institutional reporting protocols:

- **Level 4 (Serious Adverse Event)**: An error causing direct patient harm that necessitates a prolonged hospital stay.
- **Level 5 (Severe Adverse Event & Reputational Harm)**: An error resulting in formal patient complaints, loss of institutional credibility, or medical malpractice litigation.
- **Sentinel Events (SE)**: Catastrophic, unexpected occurrences involving death, permanent harm, or severe temporary harm.

Per BDMS ERMs policy, Level 4–5 incidents are reported via Occurrence Reports within 7 days, and Sentinel Events within 24 hours. Medical malpractice claims are captured via the BDMS Claim system within 7 days, in alignment with Joint Commission International (JCI) standards.[17]

Procedure Selection Rationale for Validation

Due to the large number of tracked procedures, the main text focuses selectively on critical "High-Volume, High-Stakes" interventions with profound clinical impact. Priority was given to procedures where incidents concurrently affect two lives (eg., Cesarean delivery) and those with steep learning curves (eg., minimally invasive surgeries such as shoulder rotator cuff repair and ACL reconstruction). Specifically, we highlighted procedures that demonstrated either statistically significant associations or prominent clinical trends in incident reduction, directly testing whether higher annual volumes mitigate risk. For full transparency, the complete datasets for all other evaluated procedures are available in Table S2.

Validation Process

Expert-defined volume thresholds from Phase 1 served as analytical cut-off points. Surgeons were categorized by their adherence to these thresholds over the three-year period:

- **High-Volume (Consistent)**: Met or exceeded the threshold in all three years.
- **Medium-Volume (Inconsistent)**: Met the threshold in 1–2 years.
- **Low-Volume (Infrequent)**: Did not meet the threshold in any year.

Statistical Analysis

Analysis was performed using SPSS version 28.0. The Shapiro–Wilk test confirmed non-normal distribution; thus, non-parametric tests were applied ($\alpha = 0.05$).

1. **Categorical Data**: Pearson's Chi-Square or Fisher's Exact Test (for expected frequencies < 5) was used to compare demographic data and binary safety responses between specialties.
2. **Group Comparisons**: The Kruskal–Wallis Test was employed to assess the "dose-response" relationship between volume tiers and incident rates, followed by Dunn's post-hoc test with Bonferroni correction for pairwise comparisons.
3. **Specialty Comparison**: The Mann–Whitney *U*-Test was used to compare median consensus scores between Orthopedic and Obstetric panels.

Results

Demographic Characteristics of the Experts

A total of 36 medical experts participated, comprising 17 orthopedists (47.2%) and 19 obstetricians (52.8%). As shown in Table 1, a significant difference in gender distribution was observed between the two specialties (p = 0.047); with the orthopedic group was entirely male (100%), whereas the obstetrics group included both males (73.7%) and females (26.3%). Regarding age and clinical seniority, the majority of experts were over 50 years old with more than 26 years of experience. No significant differences were found in age groups (p = 0.631) or experience levels (p = 0.617) between the fields, indicating a comparable professional seniority and expertise between both panels.

Table 1 Demographic Characteristics of the Experts

Characteristics	Orthopedists (n=17)	Obstetricians (n=19)	P-value
Gender			0.047*
Male	17 (100%)	14 (73.7%)	
Female	0 (0%)	5 (26.3%)	
Age Group			0.631
41–50 yrs	8 (47.1%)	6 (31.6%)	
51–60 yrs	6 (35.3%)	9 (47.4%)	
>60 yrs	3 (17.6%)	4 (21.0%)	

(*Continued*)

Table 1 (Continued).

Characteristics	Orthopedists (n=17)	Obstetricians (n=19)	P-value
Experience Group			0.617
11–15 yrs	2 (11.8%)	2 (10.5%)	
16–20	5 (29.4%)	3 (15.8%)	
21–25	2 (11.8%)	2 (10.5%)	
26–30	4 (23.5%)	3 (15.8%)	
>30	4 (23.5%)	9 (47.4%)	

Notes: Data are presented as n (%). Pearson's Chi-square test was applied; Fisher's exact test was used when expected cell frequencies were < 5. *p < 0.05 indicates statistical significance.

Comparison of Competency Factors Between Specialties

The analysis of 20 competency factors (C1–C20) revealed a high degree of consensus. Both groups consistently prioritized "Knowledge" (C1), "Skill" (C2), and "Attitude" (C6) as the most critical elements of competency (Median = 5, Table 2). While most factors showed no statistical variance, obstetricians displayed a higher median emphasis on "Medical Equipment/Technology" (C8) and "Communication" (C19) (Median = 5 for both) compared to orthopedists (Median = 4), reflecting the technology-intensive and multidisciplinary nature of contemporary obstetric care.

Table 2 Competency Factors Between Orthopedics and Obstetrics

Code	Competency Factors	Orthopedics (n=17) Median (IQR)	Obstetrics (n=19) Median (IQR)	P-value
C1	Knowledge	5 (1)	5 (0)	0.103
C2	Skill	5 (1)	5 (0)	0.416
C3	General CME	4 (0)	4 (0)	0.259
C4	MOC or Specific Training	4 (0)	4 (0)	0.572
C5	Experience	4 (1)	5 (0.5)	0.051
C6	Attitude	5 (1)	5 (1)	0.970
C7	Support Team	4 (1)	4 (1)	0.708
C8	Medical Equipment or Medical Technology	4 (1)	5 (1)	0.299
C9	Infrastructure	4 (0)	4 (1)	0.821
C10	Work Environment	4 (1)	4 (1)	0.810
C11	Pre-Operative Consultation	4 (0)	4 (0.5)	0.748
C12	Case Review	4 (1)	4 (0.5)	0.913
C13	Preceptor	4 (1)	4 (1)	0.427
C14	Age and chronic diseases of doctors	4 (1)	4 (1)	* 0.035
C15	Affiliation with Government Hospital	3 (1)	3 (1)	0.776
C16	Consistency in Procedures	4 (1)	4 (1)	0.590
C17	Co-Surgeon	4 (0)	4 (1)	0.090

(*Continued*)

https://doi.org/10.2147/RMHP.S608778

Table 2 (Continued).

Code	Competency Factors	Orthopedics (n=17) Median (IQR)	Obstetrics (n=19) Median (IQR)	P-value
C18	Peer Relationship	4 (1)	4 (1)	0.557
C19	Communication	4 (1)	5 (1)	0.327
C20	Budget	4 (0)	4 (0.5)	0.154

Notes: Mann–Whitney *U*-Test was applied. * $p < 0.05$ indicates statistical significance.
Abbreviations: IQR: Interquartile Range; MOC: Maintenance of Certification.

A significant difference was observed in C14 (Age and chronic diseases of doctors, p = 0.035), indicating a divergence in how each specialty perceives the impact of a surgeon's physical health on performance. Additionally, "Experience" (C5) demonstrated a borderline trend (p = 0.051), with obstetricians attributing higher importance to clinical experience. Conversely, "Affiliation with Government Hospital" (C15) was identified as the least critical factor by both groups (Median = 3), suggesting that institutional type has minimal perceived impact on individual surgical competency.

Consensus on Minimum Annual Procedural Volume

The expert consensus on procedural volume and surgical pairing requirements derived from the Delphi process is presented in Table 3. The results are categorized into three strategic safety frameworks based on procedural characteristics and clinical risk profiles:

1. **High-Volume / High-Stakes Framework**: The panel established a recommended safety threshold of 10 cases per year for fundamental procedures such as anterior cruciate ligament (ACL) reconstruction, primary total knee arthroplasty

Table 3 Consolidated Expert Consensus on Procedural Volume and Surgical Pairing Standards

Classification for Orthopedics & OB-GYN	Recommended Minimum Volume (Case/ Year)	Need Co-Surgeon (Median Score)	IQR
3.1 High Volume/High Stakes			
ACL Reconstruction	10	4	0
Primary Total Knee	10	4	1
Fracture Distal Radius	10	4	1
Fracture Proximal Humerus	6	4	1
Shoulder Rotator Cuff Repair	5	4	1
Cesarean Delivery	10	4	1.5
3.2 High-Risk/High Complexity/ Low Volume			
Fracture Acetabulum/ Pelvis	2	5	0
Revision Total Hip/Knee	1	5	0-1
Laminectomy or PLIF	4	5	1
Microscopic Discectomy	7	5	1
Cervical Fusion	4	5	1

(*Continued*)

Table 3 (Continued).

Classification for Orthopedics & OB-GYN	Recommended Minimum Volume (Case/ Year)	Need Co-Surgeon (Median Score)	IQR
3.3 Technology vs. Conventional			
Laparoscopic Hysterectomy	5	5	1
Abdominal Hysterectomy	6	4	1

Notes: Co-surgeon necessity was rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Bold text indicates classification categories of procedures. For comprehensive results of additional procedures, please refer to Table S1.
Abbreviations: OB-GYN, Obstetrics – Gynecology; ACL, anterior cruciate ligament; PLIF, posterior lumbar interbody fusion; IQR, interquartile range.

(TKA), distal radius fractures, and cesarean delivery. A calibrated threshold was determined for upper extremity and joint-preserving procedures, specifically 6 cases for proximal humerus fractures and 5 cases for shoulder rotator cuff repair. These procedures were identified as primarily volume driven. Experts assigned a consistent median co-surgeon necessity score of 4.0, suggesting that individual procedural frequency—rather than mandatory dual-expert redundancy—is the primary mechanism for maintaining proficiency and ensuring patient safety.

2. **High-Risk / High-Complexity / Low-Volume Framework**: For rare or highly complex interventions, the strategy shifted from numerical frequency to mandatory system-based redundancy. Procedures such as pelvic/acetabular fractures and revision arthroplasty had lower recommended volumes (1–2 cases/year) but were coupled with a maximum Co-Surgeon Median Score of 5.0. This signifies a mandatory requirement for dual-expert presence to mitigate risk. Technical complexity in spinal surgeries (eg., posterior lumbar interbody fusion [PLIF], cervical fusion) also necessitated high-level intraoperative support (Median Score = 5.0) regardless of volume.
3. **Technology vs. Conventional Framework**: A distinct shift was observed between laparoscopic and conventional approaches. While laparoscopic hysterectomy required a slightly lower annual volume (5 cases) than the abdominal approach (6 cases), it demanded significantly higher intraoperative redundancy (Co-surgeon Score = 5.0 vs. 4.0). This underscores the expert consensus that as surgical technology increases in complexity, team-based safeguards must be augmented to ensure patient safety.

Additional granular data supporting the findings in Table 3, including detailed expert scoring distributions and a broader list of assessed procedures, are provided in Table S1.

Clinical Incident Rates and Procedural Volume Tiers

To validate the expert-defined competency standards, clinical incident rates were analyzed across the recommended volume thresholds. Table 4 presents the statistical validation of clinical incident rates for the selected high-stakes

Table 4 Statistical Validation of Clinical Incident Rates Across Recommended Volume Thresholds for Representative Procedures (2022–2024)

Specialty	Procedure	Total (N)	Completion Status	n (Mean Occ rate ± SD)	P-value
Orthopedics	Shoulder Rotator Cuff Repair	82	Overall Status		***<0.001[a]
			Complete	13 (1.32%±2.56%)	
			Incomplete	69 (1.45%±12.04%)	
			Years of Completion		***<0.001[b]
			High	7 (1.31%±2.06%)	
			Medium	19 (0.42%±1.84%)	
			Low	56 (1.79%±13.36%)	

(*Continued*)

Table 4 (Continued).

Specialty	Procedure	Total (N)	Completion Status	n (Mean Occ rate ± SD)	P-value
	Fracture proximal humerus	478	Overall Status		
			Complete	85 (0.07%±0.49%)	*0.027[a]
			Incomplete	393 (0.25%±5.04%)	
			Years of Completion		
			High	33 (0.06%±0.32%)	0.160[b]
			Medium	116 (0.04%±0.39%)	
			Low	329 (0.30%±5.51%)	
	Primary Total Knee	153	Overall Status		0.863 [a]
			Complete	16 (0.76%±2.21%)	
			Incomplete	137 (7.54%±22.03%)	
			Years of Completion		0.281[b]
			High	12 (1.01%±2.53%)	
			Medium	17 (2.45%±3.68%)	
			Low	124 (7.99%±23.08%)	
OB-GYN	Cesarean Delivery	440	Overall Status		
			Complete	168 (0.07%±0.33%)	***<0.001[a]
			Incomplete	272 (0.07%±1.10%)	
			Years of Completion		
			High	78 (0.01%±0.11%)	***<0.001[b]
			Medium	143 (0.08%±0.35%)	
			Low	219 (0.08±1.23%)	

Notes: [a]mann–Whitney *U*-Test; [b]Kruskal–Wallis Test. Significant levels: * p < 0.05, *** p < 0.001. For the full list of procedures and complete analysis, please refer to Table S2.
Abbreviations: OB-GYN, Obstetrics - Gynecology; Occ Rate, Occurrence Rate of adverse clinical incidents.

representative procedures (as rationalized in the Methods section). Statistical analysis confirmed that attaining the recommended annual volumes significantly mitigated surgical risks. For fracture of the proximal humerus ($p = 0.027$) and shoulder rotator cuff repair ($p < 0.001$), surgeons meeting the respective thresholds demonstrated significantly lower incident rates compared to the incomplete group.

For cesarean delivery, the analysis revealed a compelling nuance ($p < 0.001$). Mathematically, the difference in mean occurrence rates between the complete and incomplete threshold groups appeared identical at first glance (0.07% ± 0.33% vs. 0.07% ± 1.10%). This baseline depression of overall incident rates across the network is driven by robust institutional safety buffers (eg, mandatory prenatal screening, standardized Clinical Practice Guidelines [CPGs], and advanced fetal monitoring) that effectively prevent generalized errors. However, high statistical significance was achieved due to the vast difference in variance distribution. The tight variance in the complete cohort demonstrates that consistently meeting the 10-case threshold serves as an essential, independent safeguard that eliminates the out-of-boundary, competency-driven adverse events found in the incomplete group.

In contrast, within the primary TKA group, surgeons meeting the 10-case threshold exhibited a markedly lower numerical incident rate compared to those who did not (0.76% ± 2.21% vs. 7.54% ± 22.03%). Although the difference did not reach statistical significance ($p = 0.863$) due to high sample variance, the descriptive clinical trend strongly aligns with the volume-outcome safety pattern seen in other high-stakes procedures.

Furthermore, evaluating performance sustainability through "Years of Completion" revealed a clear correlation between consistency and safety. Surgeons with a high frequency of meeting annual thresholds—particularly in cesarean delivery and shoulder rotator cuff repair—consistently maintained the lowest incident rates ($p < 0.001$). This suggests that long-term adherence to these safety tiers is a more robust indicator of lower clinical risk than isolated annual volume. Detailed statistical validation and sensitivity analyses are provided in Table S2.

Table 5 Expert Opinions Regarding Alternative Safety Measures for Limited-Resource Settings

Code	Opinion	Orthopedics	OB-GYN	P-value
O1	At least one assistant surgeon has performed the required number of procedures.	76.47%	89.47%	0.391
O2	A consultation has been made with a specialist or a specialized team in that field.	70.59%	63.16%	0.732
O3	A preceptor participated in the surgery.	52.94%	31.58%	0.495
O4	A pre-operative conference was conducted.	47.06%	31.58%	0.311
O5	There were assistant surgeons (paired up, but none had reached the required number of procedures).	35.29%	47.37%	0.516

Notes: Data are presented as n (%). Pearson's Chi-square test was applied; Fisher's exact test was utilized for cell frequencies < 5. Significant at p < 0.05.

Expert Opinions on Alternative Safety Measures in Limited-Resource Settings

In scenarios where the primary surgeon has not achieved the recommended annual procedural volume, the experts evaluated five alternative safety measures (Table 5). Statistical analysis revealed no significant differences between orthopedists and obstetricians across all proposed measures ($p > 0.05$), indicating a robust, shared perspective on safety protocols between these specialties.

The most highly endorsed measure by both groups was O1: Ensuring at least one assistant surgeon has performed the required number of procedures, with 89.47% agreement from obstetricians and 76.47% from orthopedists ($p = 0.391$). This was followed by O2: Consultation with a specialist or specialized team, supported by 70.59% of orthopedists and 63.16% of obstetricians ($p = 0.732$). Conversely, the pairing of assistant surgeons who had not met the volume thresholds (O5) received the lowest level of endorsement, underscoring the necessity of having at least one high-volume practitioner within the surgical team to maintain safety standards.

Discussion

Demographic Characteristics and Expert Consistency

The demographic analysis of the 36 experts (Table 1) reveals a cohort with substantial clinical seniority; the majority were over 50 years of age with more than 26 years of experience. The lack of significant differences between orthopedists and obstetricians regarding age ($p = 0.631$) or years of practice ($p = 0.617$) establishes a comparable level of seniority. This comparability is a critical factor in Delphi studies to ensure consensus stability and reliability.[16] This extensive experience ensures that the proposed thresholds (Table 3) and safety protocols (Table 5) are grounded in long-term clinical wisdom and mastery learning rather than transient practice patterns.[18] While the gender distribution difference ($p = 0.047$) reflects the historical male dominance in orthopedic surgery, this variable did not skew professional judgment. The strong consensus across competency factors suggests that surgical safety priorities are governed by standardized clinical education and frameworks like Canadian Medical Education Directions for Specialists (CanMEDS),[4,11] rather than gender-specific perspectives.[19]

Core Competency Factors and Professional Physical Readiness

The analysis of 20 competency factors (C1–C20) revealed a robust consensus. Both groups rated Knowledge (C1), Skill (C2), and Attitude (C6) at the highest level (Median = 5, Table 2), aligning with the principles of Competency-Based Medical Education (CBME) and ACGME standards, where professional competence integrates technical dexterity with clinical reasoning.[11,20–22] Interestingly, the high value placed on Communication (C19) reflects a recognition of non-technical skills as essential components of surgical mastery, aligning with international Entrustable Professional Activities (EPAs)[6,7] and curricula in Canada and Europe.[4,12] A pivotal finding is the statistical divergence regarding C14 (Age and chronic diseases of doctors, $p = 0.035$). While technical skill is a traditional focus,[23] there is increasing recognition that a surgeon's physical and cognitive readiness is vital to mastery learning and skill maintenance.[18,24] This suggests that institutions should adopt a holistic competency evaluation that considers individual physical readiness alongside procedural volume.

Clinical Significance of Minimum Volume Thresholds

Establishing a consensus minimum annual threshold of 10 cases for high-volume procedures (eg, ACL reconstruction, cesarean delivery, and distal radius fractures) serves as a critical pedagogical framework for procedural maintenance (Table 3) According to Sawyer et al,[1] procedural skills must be proven and maintained through continuous practice to prevent skill decay, a primary driver of preventable surgical errors.[24]

The recommended 10-case threshold for primary TKA did not reach statistical significance ($p = 0.863$) in clinical validation. This constraint was driven by high intra-group variance and a small sample size ($N = 16$), which masked a prominent descriptive trend where high-volume surgeons still maintained a markedly lower numerical incident rate (0.76% vs. 7.54%). This lack of significance may also suggest a “safety ceiling” effect within the network, where highly standardized perioperative protocols and mature clinical pathways (adhering to JCI standards)[17] mitigate risks typically associated with lower individual surgeon volumes.[25,26] Furthermore, the higher requirement for a co-surgeon in laparoscopic hysterectomy (Median = 5) versus the conventional approach highlights the technical complexity of minimally invasive surgery, where a skilled assistant acts as a necessary safety redundancy.[27–29]

Validation of Safety via Clinical Incident Rates

The quantitative validation (Table 4) reinforces the well-established volume-outcome relationship.[25,30,31] The finding that meeting recommended volumes significantly reduces incident rates ($p < 0.05$) underscores the necessity for monitoring volumes as a determinant of clinical quality and EPAs.[6] This is remarkably visible in cesarean delivery ($p < 0.001$), where despite identical baseline mean incident rates (0.07%) caused by robust institutional safety networks, the complete threshold group exhibited a significantly tighter variance distribution. This proves that volume targets act as an essential safeguard against competency-driven adverse events. Furthermore, evaluating performance sustainability through “Years of Completion” revealed a clear correlation between consistency and safety. Surgeons consistently meeting thresholds over multiple consecutive years demonstrated the lowest incident rates ($p < 0.001$), suggesting that the protective effect of surgical volume is cumulative and longitudinal. This supports frameworks like the EBCOG PACT curriculum,[15] which advocate for sustained competency-based assessments. In cases where statistics were not significant, the “low event rate” phenomenon likely applied; in high-safety environments, capturing rare adverse events often requires larger, multi-center datasets.[26,32,33] To optimize data-driven clinical privileging, integrating modern digital frameworks—such as large language models (LLMs), prompt engineering, and automated standard-risk knowledge graphs—can revolutionize how hospital executives track compliance and dynamically flag skill decay before incidents occur.[34,35]

Team-Based Safety as a Mitigating Factor (O1–O5)

In resource-limited settings, the endorsement of O1 (Experienced Assistant, 76–89%) and O2 (Specialist Consultation) as primary safety alternatives (Table 5) reflects a shift toward collaborative practice models. To achieve institutional safety, clinical privileging must be anchored in continuous staff evaluation, targeted knowledge assessment, and team-based support.[17,36] Furthermore, although support for O4 Pre-operative Conferences was lower among obstetricians, international standards demonstrate that structured briefings, strict checklist adherence, and closed-loop surgical safety audits are essential for reducing preventable clinical errors, regardless of individual procedural volume.[28,29,37,38]

Limitations of the Study

While this study provides a validated framework for clinical privileging, several limitations must be acknowledged:

1. Single Healthcare Network Context: All clinical validation data were gathered from a single large-scale private hospital network enforcing strict JCI guidelines. The findings may lack direct generalizability to public healthcare systems, teaching hospitals, or rural clinics operating under different resource constraints.
2. Retrospective Design: The validation of incident rates relied entirely on a retrospective review of existing hospital records (2022–2024), restricting the ability to control for all confounding clinical variables, such as varying patient comorbidities or baseline experience of the assisting staff.

3. Potential Underreporting of Incidents: Despite a structured, non-punitive reporting culture within the network, retrospective safety audits carry an inherent risk of minor clinical incident underreporting, which could artificially depress observed error rates.

Conclusion

This study establishes a comprehensive, consensus-based framework for surgical competency, validating Knowledge, Skill, and Attitude as the foundational pillars of professional performance across Orthopedics and OB-GYN. A critical finding is the recommended safety threshold of 10 cases per year for high-volume procedures (eg, cesarean delivery, ACL reconstruction), which quantitatively demonstrated a significant reduction in clinical incident rates ($p < 005$). For high-complexity or low-volume interventions, the paradigm shifts from individual numerical frequency to system-based redundancy, mandating the intraoperative presence of a qualified co-surgeon (Median Score = 5.0). Furthermore, our longitudinal data demonstrates that multi-year consistency in meeting these volume tiers is a more robust predictor of patient safety than isolated annual performance.

From a policy perspective, these findings provide a strategic risk-management roadmap for healthcare administrators. Instead of enforcing rigid, punitive volume mandates that could trigger workforce shortages or restrict healthcare access in limited-resource settings, policymakers should utilize these thresholds as dynamic “clinical triggers.” When a primary surgeon falls below the recommended benchmark, institutions can implement highly endorsed collaborative alternatives, such as pairing them with an experienced assistant surgeon, thereby preserving both operational capacity and patient safety.

Ultimately, embedding these threshold matrices into automated institutional credentialing and Enterprise Resource Planning (ERP) systems allows medical directors to transition from reactive incident investigations to proactive, data-driven quality assurance. Future research should focus on multi-center studies across diverse public and rural healthcare sectors to evaluate the generalizability of these benchmarks, while exploring the integration of real-time digital tracking systems to continuously monitor technical skill maintenance and professional readiness.

Data Sharing Statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions related to clinical incident records and hospital confidentiality policies.

Ethics Approval and Informed Consent

This study was approved by the Institutional Review Board of Bangkok Hospital Headquarters (IRB No. 2024-05-20), which provides ethical oversight for the Bangkok Dusit Medical Services (BDMS) network. The research was conducted in two phases. For the expert consultation phase (Delphi technique), all participants were provided with an information sheet detailing the study’s objectives and procedures; written informed consent was obtained from each expert prior to their participation. For the phase involving the BDMS centralized database, formal institutional authorization was granted by the Chief Executive Officer (CEO) and the respective Hospital Directors. Data access was conducted in strict accordance with the corporate Privacy Policy, which informs medical personnel regarding the use of data for research and quality improvement. All retrospective data were fully anonymized and analyzed in an aggregate format to ensure the confidentiality of all subjects.

Acknowledgments

The authors wish to express their deepest gratitude to the expert panel, who contributed their invaluable expertise to this study. We also extend our sincere thanks to the hospital directors, chief medical officers, and medical affairs officers across the participating healthcare network for their administrative support and assistance in clinical data coordination.

Funding

This research was supported and funded by Bangkok Dusit Medical Services PCL (BDMS).

Disclosure

The authors report no conflict of interest in this study.

References

1. Sawyer T, White M, Zaveri P, et al. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med*. 2015;90(8):1025–1033. doi:10.1097/ACM.0000000000000734
2. Medical Board of Australia. *Good Medical Practice: A Code of Conduct for Doctors in Australia*. Melbourne: Medical Board of Australia; 2020.
3. General Medical Council. *Outcomes for Graduates*. London: General Medical Council; 2018.
4. Frank JR, Danoff D. The CanMEDS initiative: implementing an outcomes-based model of physician competence. *Med Teach*. 2007;29(7):642–647. doi:10.1080/01421590701746983
5. Englander R, Cameron T, Ballard AJ, et al. Toward a common taxonomy of competency domains for the health professions and competencies for physicians. *Acad Med*. 2013;88(8):1088–1094. doi:10.1097/ACM.0b013e31829a3b2b
6. Ten Cate O. Competency-based education, entrustable professional activities, and the power of language. *J Grad Med Educ*. 2013;5(1):6–7. doi:10.4300/JGME-D-12-00381.1
7. Association of American Medical Colleges. *Core Entrustable Professional Activities for Entering Residency: Curriculum Developers' Guide*. Washington, DC: AAMC; 2014.
8. Accreditation Council for Graduate Medical Education. ACGME program requirements for graduate medical education in orthopedic surgery / obstetrics and gynecology. *ACGME Case Log System*. 2023.
9. Nasca TJ, Philibert I, Cassel CK, Day SH. The next GME accreditation system—from process to outcomes. *N Engl J Med*. 2012;366 (11):1051–1056. doi:10.1056/NEJMsr1200117
10. Swing SR. The ACGME outcome project: retrospective and prospective. *Med Teach*. 2007;29(7):648–654. doi:10.1080/01421590701392903
11. Frank JR, Snell LS, Cate OT, et al. Competency-based medical education: theory to practice. *Med Teach*. 2010;32(8):638–645. doi:10.3109/0142159x.2010.501190
12. Ten Cate O. Entrustability of professional activities and competency-based training. *Med Educ*. 2005;39(12):1176–1177. doi:10.1111/j.1365-2929.2005.02341.x
13. German Medical Association (Bundesärztekammer). *(Muster-)Weiterbildungsordnung 2018 (MWBO)*. Berlin: Bundesärztekammer; 2018.
14. Swiss Institute for Medical Education (SIWF). *Regulations for Further Medical Training (WBO)*. Bern: SIWF; 2022.
15. European Board and College of Obstetrics and Gynaecology (EBCOG). *The PACT Curriculum: Providing Postgraduate Academic Certified Training*. EBCOG Standards of Care and Training; 2021.
16. Macmillan TT. *The Delphi Technique*. Washington, D.C: Distributed by ERIC Clearinghouse; 1971.
17. Joint Commission International. *Joint Commission International Accreditation Standards for Hospitals*. 8th ed. Oak Brook, IL: Joint Commission Resources; 2024.
18. McGaghie WC. Mastery learning: it is time for medical education to join the 21st century. *Acad Med*. 2015;90(11):1438–1441. doi:10.1097/ACM.0000000000000911
19. Frank JR, Mungroo R, Ahmad Y, et al. Toward a definition of competency-based education in medicine: a systematic review of published definitions. *Med Teach*. 2010;32(8):631–637. doi:10.3109/0142159X.2010.500898
20. Tariq M, Bhulani N, Jafferani A, et al. Optimum number of procedures required to achieve procedural skills competency in internal medicine residents. *BMC Med Educ*. 2015;15:179. doi:10.1186/s12909-015-0464-1
21. Accreditation Council for Graduate Medical Education. Orthopedic Minimum Numbers. Available from: https://www.acgme.org/globalassets/pdfs/orthopaedic-surgery-case-log-minimums-october-2024.pdf. Accessed March 1, 2026.
22. Accreditation Council for Graduate Medical Education. Case log information: obstetrics and gynecology. Available from: https://www.acgme.org/globalassets/pfassets/programresources/obgyncaseloginfo.pdf. Accessed March 1, 2026.
23. Memon MA, Brigden D, Subramanya MS, Memon B. Assessing the surgeon's technical skills: analysis of the available tools. *Acad Med*. 2010;85 (5):869–880. doi:10.1097/ACM.0b013e3181d74216
24. Riegel B, Nafziger SD, McBurnie MA, et al. How well are cardiopulmonary resuscitation and automated external defibrillator skills retained over time? Results from the Public Access Defibrillation (PAD) trial. *Acad Emerg Med*. 2006;13(3):254–263. doi:10.1111/j.1553-2712.2006.tb00325.x
25. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med*. 2002;137(6):511–520. doi:10.7326/0003-4819-137-6-200209170-00012
26. Nallamothu BK, Krumholz HM, Ko DT, et al. Volume, quality and outcome in coronary interventions: 20 years of data and what have we learned? *J Am Coll Cardiol*. 2007;50(6):481–489. doi:10.1016/j.jacc.2007.04.045
27. Molloy D, Di-Giulio AR, Reidy J, et al. The "Surgical assistant" in laparoscopic surgery: a consensus on the role and requirements. *J Minim Invasive Gynecol*. 2006;13(6):S51. doi:10.1016/j.jmig.2006.07.113
28. World Health Organization. *WHO Guidelines for Safe Surgery 2009: Safe Surgery Saves Lives*. Geneva: World Health Organization; 2009.
29. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360(5):491–499. doi:10.1056/NEJMsa0810119
30. Wiegers EJA, Sewalt CA, Venema E, et al. The volume-outcome relationship for Hip fractures: a systematic review and meta-analysis of 2,023,469 patients. *Acta Orthop*. 2019;90(1):26–32. doi:10.1080/17453674.2018.1545383
31. Hentschker C, Mennicken R. The volume-outcome relationship revisited: practice indeed makes perfect. *Health Serv Res*. 2018;53(1):15–34. doi:10.1111/1475-6773.12696
32. Button KS, Ioannidis JP, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365–376. doi:10.1038/nrn3475
33. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305–307. doi:10.1038/d41586-019-00857-9

34. Zhu W, Zhang P, Xia W, et al. AI-Driven medical device risk management: a new paradigm integrating large language models and prompt engineering for standard-risk knowledge graph construction and application. *Risk Manag Healthc Policy*. 2026;19:345–358. doi:10.2147/RMHP.S451156
35. Ouda E, Chaabi I, Abualola H, et al. The role of artificial intelligence in reducing dispensing errors for patient safety and quality: a systems approach. *Risk Manag Healthc Policy*. 2026;19:412–425. doi:10.2147/RMHP.S453762
36. Mosalli R, Aboumoustafa GA, Khayyat W, et al. Assessment of nurses' knowledge and confidence regarding tracheostomy care in a pediatric long term care hospital in Saudi Arabia. *Risk Manag Healthc Policy*. 2022;15:1811–1821. doi:10.2147/RMHP.S374730
37. Yusuf FY, Hussein AM, Elmi AH, et al. Improving surgical safety in somalia: a closed-loop audit study of WHO surgical safety checklist adherence. *Risk Manag Healthc Policy*. 2024;17:1545–1555. doi:10.2147/RMHP.S472567
38. Srivastava M, Sadaf, Sharma SK, et al. Knowledge attitude and practices regarding surgical safety checklist: a multicenter study. *J Pharm Bioallied Sci*. 2025;17(Suppl 2):S1674–S1676. doi:10.4103/jpbs.jpbs_244_24