Cureus
Part of SPRINGER NATURE

Open Access Original Article

Development of a Consensus-Driven Checklist for Failed First-Pass Intubation Documentation in Prehospital Care: A Modified Delphi Study

Review began 10/04/2026
Review ended 10/24/2026
Published 10/25/2026

© Copyright 2026
Button et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118552

Dale Button [1, 2], Ed Lewandowski [2], Ryan Cichowski [2]

1. Paramedicine, Durham College, Oshawa, CAN 2. Central East Prehospital Care Program, Lakeridge Health, Oshawa, CAN

Corresponding author: Dale Button, dbutton@lh.ca

Abstract

Background

First-pass success is an important airway outcome, but it does not explain why an attempt failed or what changed afterward. Meaningful quality-improvement review requires documentation of patient factors, preparation, equipment, procedural barriers, corrective actions, and rescue strategy. However, a focused chart-audit rubric for failed first-pass prehospital intubation has not been established.

Objective

The study's objective is to develop an expert-consensus-derived rubric that defines the documentation content required to reconstruct and review a failed first-pass prehospital intubation attempt.

Materials and methods

We conducted a three-round electronic modified Delphi study in Ontario, Canada. Eligibility was screened before access and required at least five years of relevant airway experience plus at least two predefined indicators of clinical, educational, chart-review, or quality-improvement expertise. Forty-eight eligible individuals were invited to rate 19 candidate domains for importance and clarity on seven-point scales. Controlled feedback and revised scoring anchors were provided between rounds. Consensus for inclusion required at least 80% rating importance as six or seven; consensus for exclusion required at least 80% rating importance as one or two. Items meeting neither criterion were revised and rerated. The study closed after Round 3 when all remaining domains were clearly understood and further unchanged voting was judged unlikely to resolve stable or feasibility-based disagreement.

Results

Forty-four of 48 invitees consented and registered. Round 1 included 32 participants, Round 2 included 24, and Round 3 included 20. Among final-round participants with linked demographic records, 15 of 19 were advanced care paramedics and 15 of 19 had at least 10 years of airway-related experience. Thirteen of 19 domains reached consensus for inclusion, six reached no consensus, and none reached consensus for exclusion. Baseline vital signs before intubation, device selection, and alternative airway strategy reached inclusion consensus in Round 3. Patient positioning and airway securement each received 75% support, one response below the 80% threshold. Mean Round 3 clarity ranged from 6.15 to 6.80.

Conclusions

The Delphi process produced a 13-domain, item-anchored rubric for evaluating whether a failed first-pass intubation record contains enough information to support meaningful retrospective review. The rubric is ready for pilot evaluation in paramedic chart review and quality-improvement programs. Field testing should now establish interrater reliability, feasibility, and score interpretation before higher-stakes use.

Categories: Emergency Medicine, Medical Education, Quality Improvement
Keywords: airway management, chart review, delphi technique, emergency medical services, first-pass success, intubation, quality improvement

Introduction

Prehospital tracheal intubation is performed in unstable patients, under time pressure, and in environments where personnel, positioning, equipment access, lighting, contamination, and patient physiology may all affect performance. First-pass success is therefore an important clinical and quality-management outcome. Additional laryngoscopic attempts are associated with more adverse events, and prehospital evidence identifies airway contamination, restricted movement, trauma, anatomical distortion, environmental conditions, and equipment problems as contributors to failure [1,2].

How to cite this article
Button D, Lewandowski E, Cichowski R (October 25, 2026) Development of a Consensus-Driven Checklist for Failed First-Pass Intubation Documentation in Prehospital Care: A Modified Delphi Study. Cureus 18(10): e118552. DOI 10.7759/cureus.118552

The first-pass outcome alone is not sufficiently explanatory for quality assurance. Failed attempts occur through different mechanisms and may be followed by suctioning, repositioning, a different device or adjunct, a change in operator, another intubation strategy, or rescue airway management [3,4]. A medical oversight program that records only success, failure, or the number of attempts can identify that an adverse procedural outcome occurred, but cannot determine whether the review should focus on patient anatomy, physiological preparation, equipment, positioning, technique, team processes, documentation design, or education. Effective quality improvement therefore depends on a record that preserves the conditions surrounding the attempt, the barrier encountered, and the response to that barrier.

The need for a reconstructable record does not imply that unaided practitioner recollection is an adequate reference standard. Written documentation created after a dynamic resuscitation is vulnerable to omission and recall error. In one comparison of prehospital intubation records with video review, agreement was poor across several seemingly objective elements, including attempt timing, difficult-airway assessment, number of attempts, bag-mask ventilation, and introducer use [5]. This limitation strengthens the case for structured documentation. A dedicated airway registry form has been shown to improve recording of difficult-airway indicators, laryngoscopic grade, confirmation, and operator details compared with generic records [6]. A structured audit rubric can therefore assess whether the chart contains the information required for review, while remaining distinct from a determination of what actually occurred or whether care was appropriate.

Existing airway reporting frameworks provide an important foundation. The Utstein-style prehospital airway template and its update define standardized variables for research and system reporting, while consensus quality indicators and the National Association of Emergency Medical Services (EMS) Physicians quality-management recommendations emphasize uniform definitions, process measurement, and linkage to education and continuous improvement [7-10]. These frameworks are broader than the specific task of reviewing a failed first attempt. To our knowledge, they do not provide an item-anchored rubric that allows a reviewer to judge whether an individual paramedic record contains absent, incomplete, or sufficiently clear documentation of the failed-first-pass episode.

A modified Delphi method was appropriate because the literature identifies clinically relevant variables but does not establish which elements should form a minimum, routinely auditable documentation standard. Delphi methodology supports structured judgment from a purposively selected panel through anonymous iterative rating, controlled feedback, predefined consensus rules, and transparent stopping criteria [11-14]. The purpose of this study was to develop an expert-consensus-derived documentation audit rubric for failed first-pass prehospital intubation. The study addressed content development and operational wording so that the resulting instrument could proceed to field testing. This study was limited to consensus-based content development; it did not evaluate interrater reliability, clinical validity, implementation effectiveness, or patient outcomes.

Materials And Methods

Study design and setting

We conducted a prospective, non-interventional, three-round electronic modified Delphi consensus study from May to August 2026. The study was led by the Central East Prehospital Care Program at Lakeridge Health in Ontario, Canada. The Delphi was modified because the initial candidate domains were generated before panel rating from a targeted literature review and existing reporting frameworks. Panelists remained anonymous to one another, while the research team retained identities for eligibility screening, individualized survey access, reminders, longitudinal linkage, and duplicate management. Reporting was informed by ACCORD, CREDES, and DELPHISTAR [12-14].

Failed first-pass intubation was defined as an episode in which tracheal tube placement was not successfully completed on the first laryngoscopic attempt and was followed by another attempt, a change in approach, or a rescue airway response. The intended output was a retrospective chart review and educational tool, not a checklist to be completed during active airway management.

Development of the candidate domains

The steering group developed a 19-domain draft rubric using a targeted review of prehospital airway literature, Utstein-style reporting variables, airway quality-management recommendations, and investigator experience in paramedic practice, education, and chart review [6-10]. The review was literature-informed rather than systematic. Candidate domains were difficult airway assessment, indication for airway intervention, patient positioning, preoxygenation strategy, baseline physiology, number of attempts, device selection, adjuncts, visualization maneuvers, complications, reasons for failure, alternative airway strategy, attempt duration, primary confirmation, secondary confirmation, post-intubation physiology, airway securement, environmental conditions, and team roles or checklist use.

Each domain was expressed as a three-level documentation criterion. A score of zero represented absent documentation, one represented incomplete or nonspecific documentation, and two represented clear or complete documentation. These anchors were refined between rounds in response to clarity ratings and

written comments.

Panel eligibility and recruitment

Potential panelists were identified through a Knowledge Resource Nomination Worksheet, direct invitation, and professional and educational networks. Eligibility screening was completed before Round 1 access was issued. Participants required at least five years of relevant clinical, educational, research, or quality-improvement experience in airway management and at least two of seven expertise indicators: a peer-reviewed airway publication or national presentation; a teaching or faculty role; leadership in clinical protocols or airway quality improvement; precepting or mentoring advanced-airway learners; participation in a structured airway debrief or morbidity and mortality review; review of at least 20 airway charts; or quality-assurance or quality-improvement experience related to paramedic education or practice. The marker set was intentionally practice-based because expertise relevant to paramedic documentation review is not limited to publication history or formal protocol leadership. Individuals with an unresolved conflict of interest, fewer than five years of relevant experience, prior involvement in drafting the rubric, or inability to complete repeated English-language surveys were excluded.

Forty-eight eligible individuals were invited, and 44 provided electronic consent and completed registration. Participation was unpaid. Identifiable names and email addresses were used for eligibility administration, individualized access, longitudinal linkage, and duplicate management. Aggregate round-level characteristics were prepared without identifiers.

Delphi rounds and controlled feedback

Round 1 was open from May 14 to June 11, 2026; Round 2 from June 19 to July 10; and Round 3 from July 24 to August 10. In Round 1, panelists reviewed all 19 domains. They rated importance, defined as whether the domain was essential for a standard review of failed first-pass documentation, and clarity, defined as how clearly the scoring criterion was expressed. Both used seven-point scales anchored at one, not at all; four, neutral; and seven, extremely. Each domain also included an optional comment field.

After each round, the steering group reviewed rating distributions and all written comments. Later-round pages displayed the previous wording, revised scoring anchors, the reason for revision, and controlled aggregate feedback when relevant. Feedback included the number of respondents, the percentage rating importance as six or seven, mean clarity, the percentage rating clarity as six or seven, and the number of comments reviewed. Individual identities and comments were not shared. Controlled feedback is a core Delphi mechanism because it allows panelists to reconsider an item in light of the group response while retaining anonymity [15,16].

Items that reached the importance threshold were removed from further importance voting. If an accepted item required wording refinement, it could undergo clarity-only review in the next round without reopening the inclusion decision. Round 2 contained 14 items, including two clarity-only items. Round 3 contained nine items that had not yet reached either importance threshold. Archived item wording, documentation scoring anchors, and revision rationales for all three rounds are provided in Appendix A.

Consensus, clarity, and stopping rules

Consensus for inclusion was defined a priori as at least 80% of valid respondents rating importance as six or seven. Consensus for exclusion required at least 80% rating importance as one or two. An item meeting neither criterion was classified as no consensus and could be revised and rerated. At the observed round sizes, inclusion required at least 26 of 32 responses in Round 1, 20 of 24 in Round 2, and 16 of 20 in Round 3. Importance decisions were not changed because an item was close to the threshold.

Clarity was evaluated separately and guided wording revision rather than inclusion. All written comments were reviewed regardless of the quantitative clarity result. Three rounds were completed. At the end of Round 3, all remaining domains were clearly understood, several showed stable subthreshold ratings, and the steering group judged that another unchanged round was unlikely to resolve disagreement about importance, routine applicability, or feasibility. Remaining domains were therefore retained as no consensus rather than reclassified as excluded.

Data management and analysis

One selected response per participant per round was analyzed. Round 2 contained 26 raw submissions from 24 participants because two participants submitted updated responses; the latest complete response was retained for each. The Round 3 database was closed and locked at 20 unique complete responses before final analysis.

Analysis was descriptive and threshold-driven. For each domain and round, we calculated the full Likert distribution, the number and percentage rating importance as six or seven, the number and percentage rating importance as one or two, median and interquartile range, mean, and clarity summaries. Wilson 95%

confidence intervals were calculated for final-round high-importance proportions as measures of precision and did not alter consensus decisions. An absolute change of less than 10 percentage points between consecutive rounds was used descriptively to identify stable support. Free-text comments were used to refine wording and interpret feasibility concerns; they were not analyzed as a formal qualitative study.

Ethics

The Lakeridge Health Research Ethics Board approved the study (REB 2025-011) on June 23, 2025. Annual approval was renewed on June 1, 2026, through June 1, 2027. Participants provided electronic informed consent. No patient data were collected.

Results

Participant flow and panel characteristics

Of 48 eligible invitees, 44 provided consent and completed registration, and 32 completed Round 1, for a response rate of 66.7%. Twenty-four completed Round 2, representing 75.0% retention, and 20 completed Round 3, representing 83.3% retention from Round 2 and 62.5% of the Round 1 panel. All Round 3 participants rated all nine domains. Figure *1* presents participant flow.

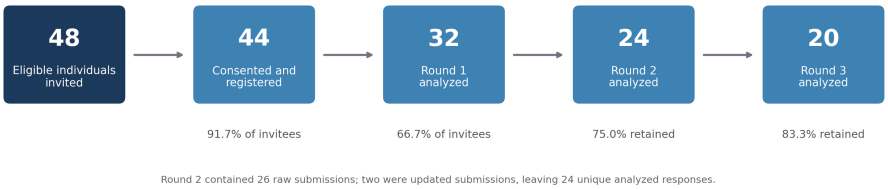


FIGURE 1: Participant Flow Across the Three Delphi Rounds

Forty-eight eligible individuals were invited, 44 consented and registered, and 32, 24, and 20 completed Rounds 1, 2, and 3, respectively. Percentages for Rounds 2 and 3 describe retention from the preceding round. Round 2 contained 26 raw submissions; two were updated submissions, leaving 24 unique analyzed responses.

Round-specific demographic records were available for 29 of 32 Round 1 participants, 23 of 24 Round 2 participants, and 19 of 20 Round 3 participants. The panel remained predominantly composed of advanced care paramedics throughout the study. Among participants with linked demographic records, advanced care paramedics accounted for 82.8% in Round 1 and 78.9% in Round 3. At least 10 years of airway-related experience was reported by 82.8% in Round 1 and 78.9% in Round 3. Teaching, precepting, and quality-improvement experience remained common across rounds, indicating that attrition did not substantially change the practice-based expertise profile of the panel. Table *1* provides round-specific characteristics.

Characteristic	Round 1	Round 2	Round 3
Demographic record available	29/32 (90.6%)	23/24 (95.8%)	19/20 (95.0%)
Primary professional role			
Advanced care paramedic	24/29 (82.8%)	18/23 (78.3%)	15/19 (78.9%)
Airway educator/faculty	3/29 (10.3%)	3/23 (13.0%)	2/19 (10.5%)
Quality-improvement lead	1/29 (3.4%)	1/23 (4.3%)	1/19 (5.3%)
Other professional role	1/29 (3.4%)	1/23 (4.3%)	1/19 (5.3%)
Airway-related experience			
5-9 years	5/29 (17.2%)	4/23 (17.4%)	4/19 (21.1%)
10-14 years	8/29 (27.6%)	5/23 (21.7%)	3/19 (15.8%)
15 years or more	16/29 (55.2%)	14/23 (60.9%)	12/19 (63.2%)
Ontario practice setting	28/29 (96.6%)	22/23 (95.7%)	19/19 (100.0%)
Selected expertise markers			
Teaching/faculty experience	26/29 (89.7%)	20/23 (87.0%)	17/19 (89.5%)
Precepting/mentoring experience	27/29 (93.1%)	21/23 (91.3%)	18/19 (94.7%)
Quality assurance/quality improvement experience	15/29 (51.7%)	11/23 (47.8%)	10/19 (52.6%)
Reviewed at least 20 airway charts	11/29 (37.9%)	9/23 (39.1%)	8/19 (42.1%)
Structured airway debrief or morbidity/mortality review	9/29 (31.0%)	6/23 (26.1%)	6/19 (31.6%)
Clinical protocol or airway-QI leadership	8/29 (27.6%)	5/23 (21.7%)	4/19 (21.1%)
Peer-reviewed publication or national presentation	2/29 (6.9%)	1/23 (4.3%)	1/19 (5.3%)

TABLE 1: Round-Specific Panel Characteristics

Professional role, experience, geography, and expertise-marker percentages use linked demographic records as the denominator. Eligibility was confirmed before Round 1 access. Expertise markers were multi-select.

QI, quality improvement

Round 1

Seven of 19 domains reached consensus for inclusion in Round 1: difficult airway assessment, number of attempts, adjuncts used, complications encountered, reasons for failure, primary confirmation, and metrics following intubation. High-importance agreement ranged from 81.2% to 90.6%. No domain reached consensus for exclusion. Difficult airway assessment and adjuncts used retained their inclusion decisions but advanced to Round 2 for clarity-only review.

Among the 12 domains without inclusion consensus, baseline physiology received 78.1% high-importance ratings, and indication, device selection, and secondary confirmation each received 75.0%. Attempt duration received 25.0%, but did not meet the exclusion threshold.

Round 2

Round 2 rerated 12 domains for importance and reviewed revised wording for two clarity-only domains. Indication for airway intervention reached 91.7% high-importance agreement, preoxygenation strategy reached 83.3%, and secondary confirmation reached 95.8%; all three achieved consensus for inclusion. Nine domains advanced to Round 3. None reached consensus for exclusion.

Round 3 and final disposition

Three of nine domains reached consensus for inclusion in Round 3: baseline vital signs before intubation at 95.0%, device selection at 90.0%, and alternative airway strategy at 85.0%. Patient positioning and securing the airway each received 15 of 20 high-importance ratings, or 75.0%. Each was one response below the 16-

of-20 inclusion requirement and remained classified as no consensus.

The remaining no-consensus domains were airway-visualization optimization maneuvers at 60.0%, environmental conditions at 50.0%, team roles and checklist use at 35.0%, and attempt duration at 15.0%. Attempt duration was not excluded because only 10.0% rated it as one or two. Across the nine Round 3 domains, mean clarity ranged from 6.15 to 6.80, and 80.0% to 100.0% rated clarity as six or seven. The remaining disagreement therefore concerned whether domains should be required in routine audit, not whether their wording was understood. Table *2* shows final disposition and Figure *2* shows the decision round for each domain. Detailed Round 3 results are shown in Table *3*.

Item	Domain	Decision round	Importance rated 6 or 7	Final disposition
1	Difficult airway assessment	1	27/32 (84.4%)	Consensus for inclusion
2	Indication for airway intervention	2	22/24 (91.7%)	Consensus for inclusion
3	Patient positioning	3	15/20 (75.0%)	No consensus
4	Preoxygenation strategy	2	20/24 (83.3%)	Consensus for inclusion
5	Baseline vital signs prior to intubation	3	19/20 (95.0%)	Consensus for inclusion
6	Number of attempts	1	26/32 (81.2%)	Consensus for inclusion
7	Device selection	3	18/20 (90.0%)	Consensus for inclusion
8	Adjuncts used	1	27/32 (84.4%)	Consensus for inclusion
9	Airway visualization optimization maneuvers	3	12/20 (60.0%)	No consensus
10	Complications encountered	1	29/32 (90.6%)	Consensus for inclusion
11	Reasons for failure	1	28/32 (87.5%)	Consensus for inclusion
12	Alternative airway strategy, if selected	3	17/20 (85.0%)	Consensus for inclusion
13	Attempt duration	3	3/20 (15.0%)	No consensus
14	Primary confirmation	1	28/32 (87.5%)	Consensus for inclusion
15	Secondary confirmation	2	23/24 (95.8%)	Consensus for inclusion
16	Metrics following intubation	1	26/32 (81.2%)	Consensus for inclusion
17	Securing the airway	3	15/20 (75.0%)	No consensus
18	Environmental conditions	3	10/20 (50.0%)	No consensus
19	Team roles and checklist use	3	7/20 (35.0%)	No consensus

TABLE 2: Final Disposition of the 19 Candidate Documentation Domains

For domains reaching consensus for inclusion, the percentage is from the decision round. For no-consensus domains, the percentage is from Round 3. Consensus for exclusion required at least 80% rating importance as one or two; no domain met that criterion.

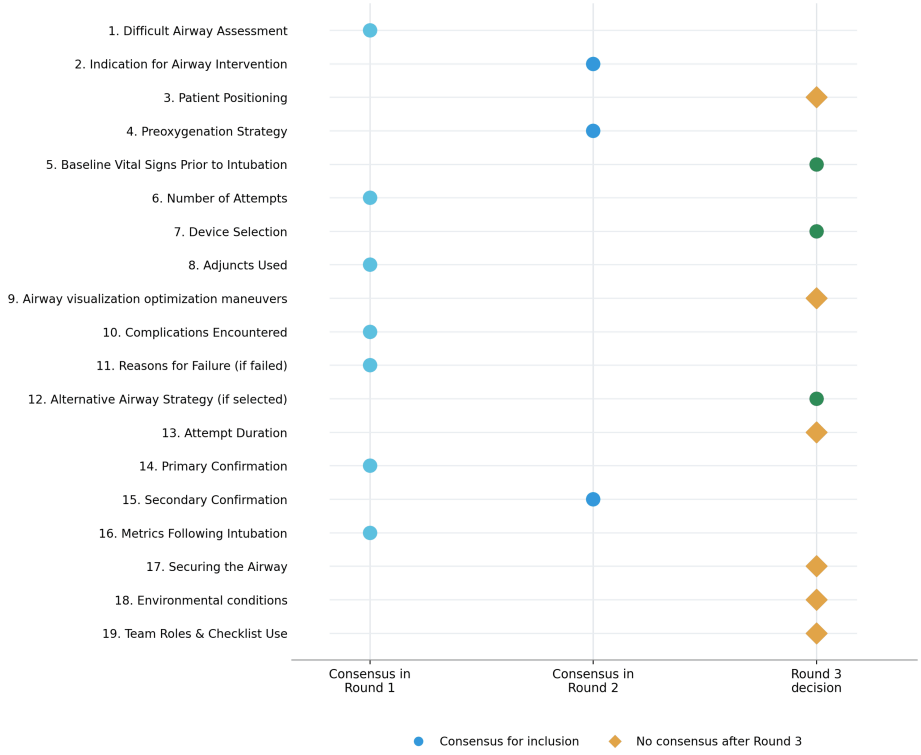


FIGURE 2: Final Disposition and Decision Round for the 19 Candidate Documentation Domains

Circles indicate the round in which a domain reached consensus for inclusion. Diamonds indicate domains with no consensus after Round 3. No domain reached consensus for exclusion.

Item	Domain	Importance 6-7	Wilson 95% CI	Median (IQR)	Clarity mean	Clarity 6-7	Disposition
3	Patient positioning	15/20 (75.0%)	53.1-88.8	6.0 (5.5-7.0)	6.65	19/20 (95.0%)	No consensus
5	Baseline vital signs prior to intubation	19/20 (95.0%)	76.4-99.1	7.0 (6.0-7.0)	6.7	19/20 (95.0%)	Consensus for inclusion
7	Device selection	18/20 (90.0%)	69.9-97.2	6.0 (6.0-7.0)	6.75	19/20 (95.0%)	Consensus for inclusion
9	Airway visualization optimization maneuvers	12/20 (60.0%)	38.7-78.1	6.0 (5.0-7.0)	6.55	19/20 (95.0%)	No consensus
12	Alternative airway strategy (if selected)	17/20 (85.0%)	64.0-94.8	6.0 (6.0-7.0)	6.75	20/20 (100.0%)	Consensus for inclusion
13	Attempt duration	3/20 (15.0%)	5.2-36.0	4.5 (3.5-5.0)	6.15	16/20 (80.0%)	No consensus
17	Securing the airway	15/20 (75.0%)	53.1-88.8	6.0 (5.5-7.0)	6.8	20/20 (100.0%)	No consensus
18	Environmental conditions (e.g., poor lighting, confined space, hostile environment)	10/20 (50.0%)	29.9-70.1	5.5 (4.0-7.0)	6.7	20/20 (100.0%)	No consensus
19	Team roles and checklist use	7/20 (35.0%)	18.1-56.7	4.0 (3.0-6.0)	6.45	19/20 (95.0%)	No consensus

TABLE 3: Detailed Round 3 Domain Results

Confidence intervals describe precision and did not determine consensus.

CI, confidence interval; IQR, interquartile range

Airway-visualization maneuvers, environmental conditions, and team roles or checklist use changed by fewer than 10 percentage points between Rounds 2 and 3 and remained below the inclusion threshold. This stable nonconsensus supported closure without interpreting the domains as unimportant or excluded.

Overall, 13 of 19 candidate domains reached consensus for inclusion, six reached no consensus, and none reached consensus for exclusion. The final 13-domain rubric is provided in Appendix B.

Comment-informed interpretation

The dataset contained 302 written comments across the three rounds, including 22 in Round 3. Comments distinguished clinical relevance from routine documentability. Attempt duration generated repeated concern about whether exact timing could be captured accurately without automated timestamps, monitor data, or video. Other comments questioned whether conditionally relevant elements, such as visualization maneuvers, environmental barriers, or team roles, should be mandatory when they did not alter the attempt. Comments also identified possible duplication with existing procedure fields and the need for clearer examples or electronic prompts. These comments informed wording and interpretation but did not override the quantitative decision rules.

Discussion

Principal findings and practical contribution

This modified Delphi study produced a 13-domain audit rubric for failed first-pass prehospital intubation documentation. The included domains span the episode from difficult-airway assessment and indication through preparation, equipment, attempt history, complications, reason for failure, rescue strategy, confirmation, and post-intubation physiology. The practical contribution is not simply a list of airway variables. Each domain is paired with explicit absent, incomplete, and clear documentation anchors that allow a reviewer to determine whether the record is sufficiently informative for quality-improvement review.

First-pass success describes a procedural outcome, but it does not explain the mechanism of failure or

identify the most appropriate system response. The resulting rubric distinguishes a chart that merely reports failure from one that supports reconstruction of the event. During pilot use, item-level findings can identify recurrent documentation gaps, guide case review and practitioner feedback, inform educational priorities, and support the design of electronic documentation prompts. The rubric can also create more consistent data for subsequent analysis of why first-pass attempts fail. These data are necessary to design continuing professional education and quality-improvement interventions that target recurring mechanisms of failure and documentation gaps.

Relationship to existing airway quality frameworks

The domains align with Utstein-style airway variables, prehospital airway quality indicators, and quality-management recommendations [7-10]. Those resources standardize broad research and program reporting. The present rubric narrows that architecture to a specific quality-assurance question: whether an individual failed-first-pass chart contains enough information to understand the conditions, actions, and decisions surrounding the attempt. It therefore complements registry and performance frameworks rather than replacing them.

The content also maps to mechanisms described in failed-intubation studies. Difficult-airway assessment, complications, reasons for failure, devices, adjuncts, and rescue strategy correspond to contamination, anatomy, positioning, and equipment challenges reported in prehospital cohorts [2-4]. Preoxygenation, baseline physiology, confirmation, and post-intubation metrics preserve the physiological context of the airway episode. Including both number of attempts and reason for failure moves review beyond counting attempts toward identifying what changed after the first attempt.

Interpretation of domains without consensus

The six no-consensus domains should not be interpreted as rejected content. Patient positioning and airway securement each fell one response below inclusion, indicating substantial support that may justify optional collection or further study. Visualization maneuvers and environmental conditions appeared condition-dependent. Team roles and checklist use raised questions about overlap with other records, while exact attempt duration raised observability and workflow concerns. These findings may inform conditional fields, automated capture, or non-scored contextual elements rather than mandatory core scoring.

The distinction between no consensus and consensus for exclusion is important in Delphi reporting. Similar modified Delphi studies use repeated online rating and predefined thresholds to preserve disagreement rather than forcing unanimity [17,18]. Retaining the six domains and their distributions provides a transparent basis for future reconsideration in different documentation systems.

Use of the rubric and next validation phase

Content definition was necessary before field testing could occur. The current study establishes the domains and operational anchors that reviewers can now apply to actual paramedic charts. The next phase should use a sample of failed first-pass records independently scored by at least two reviewers. Testing should measure item-level agreement, weighted kappa where appropriate, review time, missingness, floor and ceiling effects, and the frequency and causes of disagreement. Reviewer cognitive debriefing can identify anchors that remain difficult to apply.

The rubric can now undergo pilot evaluation in quality-assurance settings, with item-level findings used for learning and with feasibility and reliability assessed before routine implementation. A summed score is mathematically possible, but equal weighting, dimensionality, measurement error, and score interpretation have not yet been established. Reliability and other measurement properties are distinct from the content-development work completed here [19]. High-stakes comparisons between practitioners or services should therefore wait until reproducibility and external performance are demonstrated.

Strengths

Strengths include predefined inclusion and exclusion thresholds, separate importance and clarity assessment, three iterative rounds, controlled feedback, transparent attrition, complete final-round ratings, stable identity linkage across rounds, and preservation of no-consensus outcomes. The broad expertise-marker framework recognized practice, teaching, precepting, chart review, debriefing, and quality improvement rather than equating expertise with academic publication alone. The final product translates conceptual domains into operational chart-review anchors.

Limitations

The panel was predominantly composed of Ontario advanced care paramedics, improving relevance to the intended initial setting but limiting transferability to other professions and jurisdictions. Demographic data were unavailable for three Round 1 participants, including one retained through Round 3, although eligibility was confirmed before survey access. The initial domain set was literature-informed rather than generated through a systematic review or an open-ended first round, so unanticipated content may have

been missed. Attrition reduced the final panel to 20, making the 80% threshold sensitive to a single response. Field testing has not yet established feasibility, interrater reliability, score behavior, external validity, or relationships with clinical outcomes. The study did not test whether use of the rubric improves documentation, clinical care, or patient safety.

Conclusions

A three-round modified Delphi process established consensus on 13 core domains for documenting failed first-pass prehospital intubation and translated those domains into explicit chart-review anchors. The rubric provides a structured method for evaluating whether a patient-care record contains enough information to reconstruct the failed attempt and support meaningful quality assurance, education, and case review. Six domains remained without consensus, and none reached consensus for exclusion.

The content-development phase is now complete. The rubric can proceed to pilot application in actual paramedic charts, where the next priorities are interrater reliability, feasibility, reviewer burden, and score interpretation. Until those properties are established, the tool is best used as an item-level learning and documentation-improvement framework rather than a high-stakes comparative score.

Appendices

Appendix A

Archived Delphi Item Wording and Revision Rationales

The archived item wording, documentation scoring anchors, and revision rationales for all three Delphi rounds are available in the supplementary document: https://docs.google.com/document/d/16u7E8UqkjFSZuBMhOjXus-fVmCUOsuzhPAfQ8zXnptc

Appendix B

Final 13-Domain Failed First-Pass Intubation Documentation Audit Rubric

The rubric is intended for pilot chart review and quality-improvement use. Initial implementation should emphasize item-level findings. A summed score should not be used for high-stakes comparison until reliability, weighting, and score interpretation have been established.

Item	Domain	0 - Absent	1 - Incomplete or nonspecific	2 - Clear or complete
1	Difficult airway assessment	No difficult airway assessment documented	Isolated difficult airway predictors documented, but no clear overall difficult airway assessment (e.g., one of: “blood in airway,” “small mouth opening,” “limited neck movement”).	Clear difficult airway assessment documented, including multiple relevant predictors or use of a structured tool (e.g., airway contamination with reduced jaw mobility and poor positioning or clearly documented LEMON, MACOCHA, etc.).
2	Indication for airway intervention	No indication for airway intervention documented.	Indication documented but nonspecific or lacking supporting clinical context (e.g., “airway compromise” without objective findings or rationale of what that might be or how it drove intubation decision).	Clear indication or rationale documented with objective findings or clinical context (e.g., hypoxemia/SpO2 concern, inability to ventilate or oxygenate, airway obstruction or trauma, failed less invasive airway management, predicted deterioration, or need for airway protection).
4	Preoxygenation strategy	No documentation of preoxygenation	Pre-oxygenation documented but method not specified or unclear (i.e., patient pre-oxygenated prior to attempt with no means mentioned).	Preoxygenation method clearly specified (e.g., non-rebreather mask, bag-valve-mask ventilation, supraglottic airway, or nasal cannula/apneic oxygenation)
5	Baseline vital signs prior to intubation	None of the obtainable baseline end-tidal carbon dioxide, BP, HR, SpO2, or GCS measures are documented prior to intubation.	At least one, but not all, obtainable baseline end-tidal carbon dioxide, BP, HR, SpO2, or GCS measures are documented prior to intubation.	All obtainable baseline end-tidal carbon dioxide, BP, HR, SpO2, and GCS measures are documented prior to intubation.
	Number of	No documentation of	Number of attempts documented	Number of attempts documented with clearly

6	attempts	attempts	but no comments surrounding change to strategy or approach	documented update to strategy and approach.
7	Device selection	No documentation of intubation equipment selection.	Documentation includes one or more key intubation equipment details but is missing at least one of the following: laryngoscope type, blade type, or endotracheal tube size.	Documentation includes all key intubation equipment details: laryngoscope type, blade type, and endotracheal tube size.
8	Adjuncts used	No documentation of airway adjuncts used.	Partial documentation of airway adjuncts used, with one or more adjuncts omitted or insufficiently specified.	All airway adjuncts used are documented with relevant details when available (e.g., bougie, stylet, oral or nasal airway).
10	Complications encountered	No documentation of complications	Some complications noted but lacks details (i.e., blood in airway rather than blood obstructing view of vocal cords)	Full documentation of any encountered complications (e.g., could not visualize cords due to airway positioning, excessive fluid obstructing view of cords, challenging to advance tube towards trachea)
11	Reasons for failure (if failed)	No documentation of reason for failure	Reason for failure noted but vague (e.g., “difficult airway”)	Detailed reasoning for failure (e.g., “could not position tongue, unable to see vocal cords, resistance to tube passage, blade size too big or tube size too small”)
12	Alternative airway strategy (if selected)	No backup or rescue airway strategy documented following a failed or unsuccessful intubation attempt	Backup or rescue airway strategy documented following a failed or unsuccessful intubation attempt, but with insufficient detail or decision-making rationale	Backup or rescue airway strategy fully documented following a failed or unsuccessful intubation attempt, including decision-making rationale
14	Primary confirmation	No documentation of end-tidal carbon dioxide confirmation	End-tidal carbon dioxide documented but no waveform mention	End-tidal carbon dioxide confirmation documented with waveform description
15	Secondary confirmation	No documentation of secondary confirmation	Secondary confirmation documented using only one secondary method or with unclear/incomplete details	Secondary confirmation fully documented with at least two secondary methods (e.g., auscultation, chest rise, misting)
16	Metrics following intubation	No documentation of post-intubation end-tidal carbon dioxide, BP, HR, SpO2	Post-intubation metrics recorded but incomplete	Full post-intubation metrics documented (end-tidal carbon dioxide, BP, HR, SpO2)

TABLE 4: Final Consensus-Derived Documentation Audit Rubric

The author-developed rubric is intended for pilot chart review and quality-improvement use. Initial implementation should emphasize item-level findings. A summed score should not be used for high-stakes comparison until reliability, weighting, and score interpretation have been established. Typographical and capitalization normalization did not alter the consensus-derived meaning of the anchors.

BP, blood pressure; GCS, Glasgow Coma Scale; HR, heart rate; SpO2, peripheral oxygen saturation

The following figures present the Round 3 consensus trajectories, the distribution of Round 3 importance ratings, and an example of controlled feedback and revised wording, respectively.

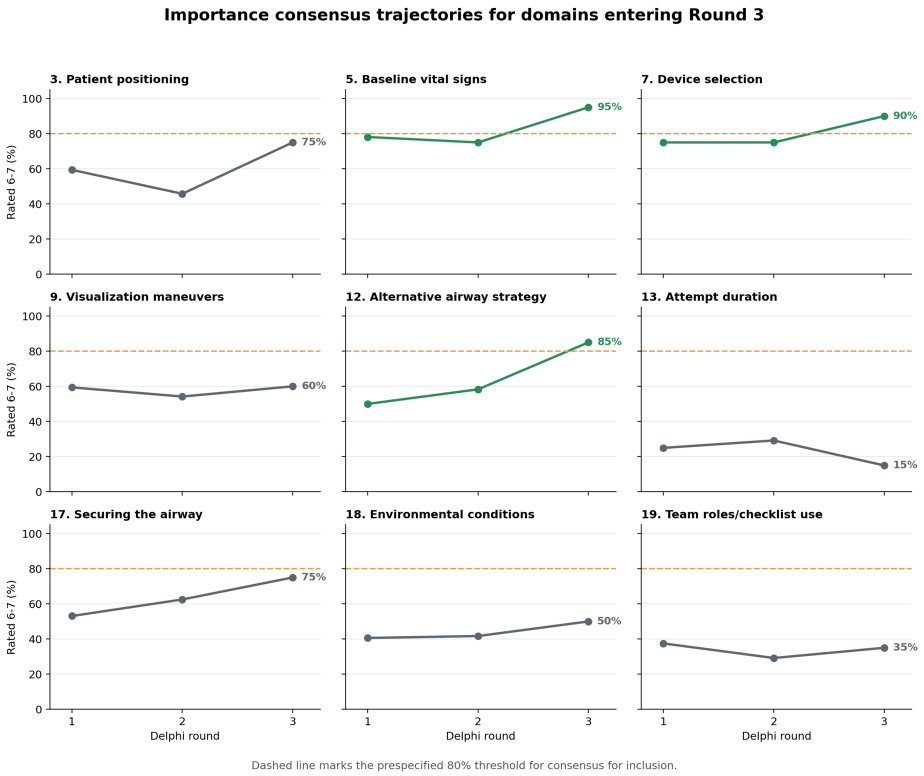


FIGURE 3: Importance Consensus Trajectories for the Nine Domains Entering Round 3

Points show the percentage rating importance as six or seven. The dashed reference line marks the prespecified 80% threshold for consensus for inclusion.

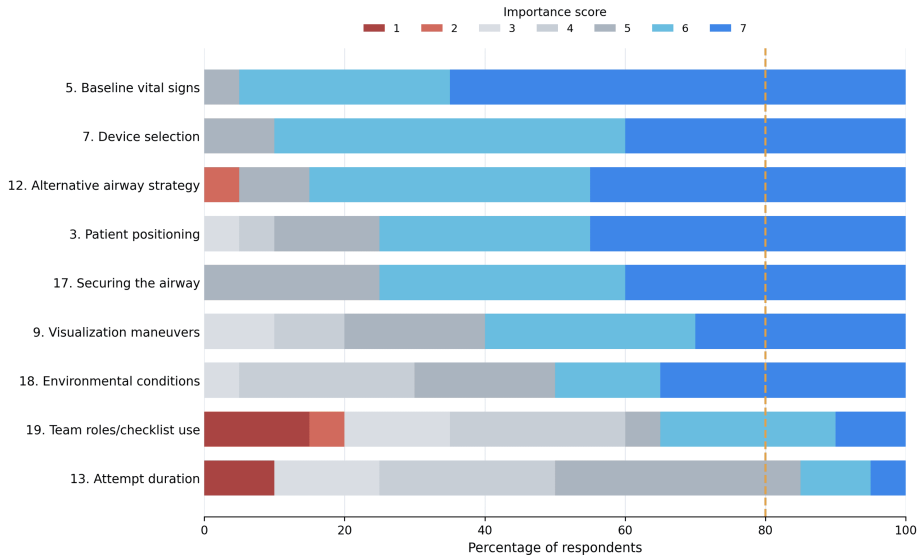


FIGURE 4: Distribution of Round 3 Importance Ratings

Domains are ordered by the percentage rating importance as six or seven. The dashed reference line marks the 80% inclusion threshold.

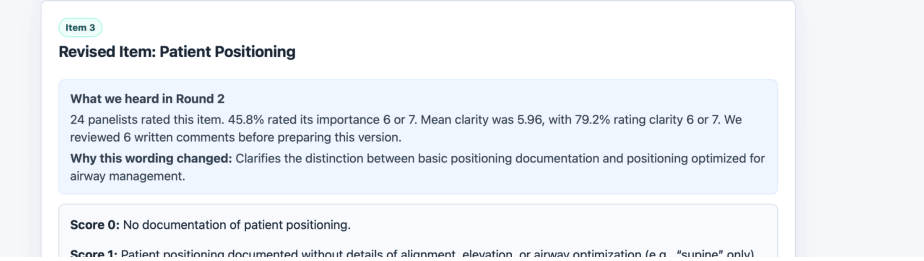


FIGURE 5: Example of Controlled Feedback and Revised Wording Provided Between Delphi Rounds

The survey page displayed the preceding-round response count, importance and clarity summaries, number of comments reviewed, and the steering group rationale for wording revision. The remainder of the scoring anchors continued below the captured image.

Clarity was evaluated separately from importance. During the first round, wording review considered both mean clarity and the percentage rating clarity as six or seven. In later rounds, mean clarity below 6.0 was the formal quantitative trigger, although all written comments were reviewed irrespective of score. This operational refinement affected wording review only and did not alter the prespecified importance thresholds or final inclusion decisions.

A change of less than 10 percentage points in high-importance agreement between consecutive rounds was labelled stable for descriptive interpretation. Stability did not itself confer inclusion or exclusion. It was considered alongside clarity and written comments when the steering group determined that another unchanged round was unlikely to add information.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Dale Button, Ed Lewandowski, Ryan Cichowski

Acquisition, analysis, or interpretation of data: Dale Button, Ed Lewandowski, Ryan Cichowski

Drafting of the manuscript: Dale Button

Critical review of the manuscript for important intellectual content: Dale Button, Ed Lewandowski, Ryan Cichowski

Supervision: Ed Lewandowski, Ryan Cichowski

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. Lakeridge Health Research Ethics Board issued approval REB 2025-011. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

The authors thank the airway clinicians, educators, and quality-improvement professionals who participated in the Delphi panel. De-identified item-level ratings, response distributions, and wording history are retained in the institutional analytical archive. They are not publicly posted because the small expert panel creates a residual risk of re-identification. Reasonable access requests will be considered under Lakeridge Health research-governance requirements. ChatGPT (OpenAI, San Francisco, CA) was used after completion of data collection and the locked analysis to assist with manuscript organization, language refinement, and reference formatting. The authors independently verified all data, analyses, interpretations, citations, and

final wording and retain full responsibility for the manuscript.

References

1. Sakles JC, Chiu S, Mosier J, Walker C, Stolz U: The importance of first pass success when performing orotracheal intubation in the emergency department. Acad Emerg Med. 2013, 20:71-8. 10.1111/acem.12055
2. Hayes-Bradley C, McCreery M, Delorenzo A, Bendall J, Lewis A, Bowles KA: Predictive and protective factors for failing first pass intubation in prehospital rapid sequence intubation: an aetiology and risk systematic review with meta-analysis. Br J Anaesth. 2024, 132:918-35. 10.1016/j.bja.2024.02.004
3. Prekker ME, Kwok H, Shin J, Carlbom D, Grabinsky A, Rea TD: The process of prehospital airway management: challenges and solutions during paramedic endotracheal intubation. Crit Care Med. 2014, 42:1372-8. 10.1097/CCM.0000000000000213
4. Gaither JB, Spaite DW, Stolz U, Ennis J, Mosier J, Sakles JJ: Prevalence of difficult airway predictors in cases of failed prehospital endotracheal intubation. J Emerg Med. 2014, 47:294-300. 10.1016/j.jemermed.2014.04.021
5. Keller CS, Dilger C, Chou SC, Lawal R, Jenks S: Video documentation as a measure of written documentation accuracy in emergency medical service field intubations. HCA Healthc J Med. 2023, 4:279-82. 10.36518/2689-0216.1183
6. Bloomer R, Burns BJ, Ware S: Improving documentation in prehospital rapid sequence intubation: investigating the use of a dedicated airway registry form. Emerg Med J. 2013, 30:324-6. 10.1136/emermed-2011-200715
7. Sollid SJ, Lockey D, Lossius HM: A consensus-based template for uniform reporting of data from pre-hospital advanced airway management. Scand J Trauma Resusc Emerg Med. 2009, 17:58. 10.1186/1757-7241-17-58
8. Sunde GA, Kottmann A, Heltne JK, et al.: Standardised data reporting from pre-hospital advanced airway management - a nominal group technique update of the Utstein-style airway template. Scand J Trauma Resusc Emerg Med. 2018, 26:46. 10.1186/s13049-018-0509-y
9. Kottmann A, Krüger AJ, Sunde GA, et al.: Establishing quality indicators for pre-hospital advanced airway management: a modified nominal group technique consensus process. Br J Anaesth. 2022, 128:e143-50. 10.1016/j.bja.2021.08.031
10. Vithalani V, Sondheim S, Cornelius A, Gonzales J, Mercer MP, Burton B, Redlener M: Quality management of prehospital airway programs: an NAEMSP position statement and resource document. Prehosp Emerg Care. 2022, 26:14-22. 10.1080/10903127.2021.1989530
11. Nasa P, Jain R, Juneja D: Delphi methodology in healthcare research: how to decide its appropriateness . World J Methodol. 2021, 11:116-29. 10.5662/wjm.v11.i4.116
12. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG: Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med. 2017, 31:684-706. 10.1177/0269216317690685
13. Gattrell WT, Logullo P, van Zuuren EJ, et al.: ACCORD (ACcurate COnsensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLoS Med. 2024, 21:e1004326. 10.1371/journal.pmed.1004326
14. Niederberger M, Schifano J, Deckert S, et al.: Delphi studies in social and health sciences—recommendations for an interdisciplinary standardized reporting (DELPHISTAR). Results of a Delphi study. PLoS One. 2024, 19:e0304651. 10.1371/journal.pone.0304651
15. Barrios M, Guilera G, Nuno L, Gomez-Benito J: Consensus in the Delphi method: what makes a decision change?. Technol Forecast Soc Change. 2021, 163:120484. 10.1016/j.techfore.2020.120484
16. Veugelers R, Gaakeer MI, Patka P, Huijsman R: Improving design choices in Delphi studies in medicine: the case of an exemplary physician multi-round panel study with 100% response. BMC Med Res Methodol. 2020, 20:156. 10.1186/s12874-020-01029-4
17. Jordans IP, Verberkt C, De Leeuw RA, et al.: Definition and sonographic reporting system for Cesarean scar pregnancy in early gestation: modified Delphi method. Ultrasound Obstet Gynecol. 2022, 59:437-49. 10.1002/uog.24815
18. Green P, Edwards EJ, Tower M: Core procedural skills competencies and the maintenance of procedural skills for medical students: a Delphi study. BMC Med Educ. 2022, 22:259. 10.1186/s12909-022-03323-9
19. Mokkink LB, Terwee CB, Patrick DL, et al.: The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010, 63:737-45. 10.1016/j.jclinepi.2010.02.006