Temporal Trends in Difficult and Failed Tracheal Intubation in a Regional Community Anesthetic Practice

Rebecca A. Schroeder, M.D., Richard Pollard, M.D., Ishwori Dhakal, M.Sc., Mary Cooter, M.Sc., Solomon Aronson, M.D., Katherine Grichnik, M.D., William Buhrman, M.D., Miklos D. Kertai, M.D., Ph.D., Joseph P. Mathew, M.D., Mark Stafford-Smith, M.D., C.M., F.R.C.P.C.

ABSTRACT

Background: When tracheal intubation is difficult or unachievable before surgery or during an emergent resuscitation, this is a critical safety event. Consensus algorithms and airway devices have been introduced in hopes of reducing such occurrences. However, evidence of improved safety in clinical practice related to their introduction is lacking. Therefore, we selected a large perioperative database spanning 2002 to 2015 to look for changes in annual rates of difficult and failed tracheal intubation.

Methods: Difficult (more than three attempts) and failed (unsuccessful, requiring awakening or surgical tracheostomy) intubation rates in patients 18 yr and older were compared between the early and late periods (pre- vs post-January 2009) and by annual rate join-point analysis. Primary findings from a large, urban hospital were compared with combined observations from 15 smaller facilities.

Results: Analysis of 421,581 procedures identified fourfold reductions in both event rates between the early and late periods (difficult: 6.6 of 1,000 vs. 1.6 of 1,000, P < 0.0001; failed: 0.2 of 1,000 vs. 0.06 of 1,000, P < 0.0001), with join-point analysis identifying two significant change points (2006, P = 0.02; 2010, P = 0.03) including a pre-2006 stable period, a steep drop between 2006 and 2010, and gradual decline after 2010. Data from 15 affiliated practices (442,428 procedures) demonstrated similar reductions.

Conclusions: In this retrospective assessment spanning 14 yr (2002 to 2015), difficult and failed intubation rates by skilled providers declined significantly at both an urban hospital and a network of smaller affiliated practices. Further investigations are required to validate these findings in other data sets and more clearly identify factors associated with their occurrence as clues to future airway management advancements.

Visual Abstract: An online visual overview is available for this article at http://links.lww.com/ALN/B635. (Anesthesiology 2018; 128:502-10)

TRACHEAL intubation is performed in a variety of settings by clinicians with a spectrum of experience who sometimes face unanticipated challenges. Routine intubation procedures involve direct laryngoscopy, during which a handheld light-tipped rigid blade is inserted into a patient’s mouth to displace the tongue and epiglottis, allowing line-of-sight vocal cord visualization for tube placement. Added preparation is typical if previous attempts have been unsuccessful (i.e., difficult) or other factors that predict difficulty are present. Risk factors generally involve variations of normal head and neck anatomy (e.g., protuberant teeth, large tongue/small pharynx, limited neck motion, obesity, previous neck radiation, history of sleep apnea, etc.).1–6 Recent changes to standard clinical practice when a difficult intubation is anticipated include guidance from expert consensus algorithms (e.g., American Society of Anesthesiologists, 1993, updated in 2003 and 2013) and availability of various advanced airway devices.2,7–11 Several advanced airway devices have been studied in patients with known difficult airway; these either offer an indirect view of the vocal cords (e.g., flexible fiberoptic scope, videolaryngoscope) or use “blind” approaches (e.g., intubating laryngeal mask airway, light-ward) to achieve tracheal intubation.2,12 Clinical trials have found many such devices to be superior to traditional direct laryngoscopy in certain settings, but their most appropriate role in routine clinical practice has not been confirmed.2,12,13

Although tracheal intubation is generally a nonemergent procedure, there is always potential for it to become a critical safety event resulting in serious harm and possibly even a fatal...
outcome. Consequently, there has been widespread incorporation of consensus algorithms and advanced devices into clinical practice to improve management of the difficult intubation. Unfortunately, looking in the existing literature for evidence of a reduction in the incidence of difficult or failed tracheal intubation events related to the addition of algorithms and devices is problematic. Published rates among cohort studies vary widely (e.g., difficult intubation: 1.5 to 16%)\(^1\); this is likely attributable in large part to the numerous definitions used for such events\(^1\) but also to the lack of conformity of other factors including practitioner experience and patient risk profiles.\(^2\)\(^\text{2,21}\) Two studies, using closed claims data spanning a period between 1975 and 2000, offer indirect evidence of a possible national decline in critical events related to airway management that overlaps the timing of relevant technical innovations (e.g., the introduction of pulse oximetry and capnography), but no equivalent more recent assessments are available, through closed claims or a patient cohort approach.\(^2\)\(^\text{2,23}\) Therefore, we used data from a perioperative quality assurance database spanning a continuous 14-yr period (2002 to 2015) to investigate difficult and failed tracheal intubation event rates over time. The database reflects anesthesia practice patterns at one large urban community hospital and 15 affiliated sites, includes standardized definitions for difficult and failed tracheal intubation, and focuses on a time period spanning the introduction of American Society of Anesthesiologists consensus guideline updates and changes in available advanced airway devices.\(^2\)\(^\text{2,8}\)

Materials and Methods

After approval by the institutional review board, deidentified data from January 2002 to September 2015 was obtained from the quality assurance database of a large regional community-based anesthesiology group practice (MEDNAX, Inc.) located in the Mid-Atlantic region of the United States.\(^24\) Patients received care at various network facilities that differ by size (outpatient-only surgery center to more than 200 inpatient beds), location (urban, suburban, rural), caseload (less than 2,000 to more than 80,000/yr), and payer mix. The largest of these sites, an urban hospital that accounted for approximately 50% of cases, was identified a priori as the primary study site.

Quantum\textsuperscript{TM} Clinical Navigation System (Q-CNS) is an internally designed quality assurance program in which data are collected prospectively for purposes of internal quality reporting. Reporting by anesthesia providers is initiated by box shading of a restricted report form (i.e., limited options with no free text), unrelated to the clinical anesthesia record. The handwritten version (fig. 1A) is scanned into a database for storage and analysis using optical character recognition, whereas the electronic version is equivalently processed through the electronic medical record (fig. 1B). Although the system remained handwritten throughout at the primary study site, completely electronic data capture was implemented at several of the secondary sites during the study period.\(^24\) Data from the entered report were subsequently validated within 48 h by a trained quality improvement (QI) nurse, any missing fields completed using both the anesthetic and electronic medical record, and maintained in a central database. At all except one of the secondary sites, Q-CNS was implemented either before or very soon after the start of the study period.

Data quality validation included an initial process occurring approximately 90 days after system deployment. In subsequent years, audits at each site were performed to evaluate for accuracy and interrater reliability. For example, data available from the largest system-wide assessment involved 705,331 cases from the period 2009 to 2014 and demonstrated an average 87.3% capture of all cases by the QI nurses during the period (Richard Pollard, M.D., Mednex, Charlotte, North Carolina; personal verbal communication, January 2017). Of these records, approximately 1.1% were sampled from the primary and each of the secondary sites for comparison against original paper or electronic health and anesthesia records to confirm QI data quality, while ensuring a minimum rate per QI nurse; an average of 181 cases per nurse were reviewed, representing 0.84% of each individual’s total caseload. Interrater reliability was calculated at 99.6±0.11% with 92% requiring at most two corrections per case, reflecting 99.6±0.16% accuracy. A review of audit activity showed that only 2% of cases required more than 2 corrections. For comparison, during the same approximate time period, the American College of Surgeons National Surgical Quality Improvement Program (https://www.facs.org/quality-programs/acs-nsqip/program-specifics/data; accessed October 25, 2017) required source electronic health and paper records and a patient phone call to review 12 to 15 records per institution annually and recommended corrective action when the disagreement rate exceeded 5%. Similarly, the Society of Thoracic Surgeons National Database (https://www.sts.org/registries-research-center/sts-national-database/sts-national-database-audits; accessed October 26, 2017) required review of 5 to 10 records per time frame, with an accuracy of 95% deemed acceptable, whereas the Centers for Medicare and Medicaid Services (https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/ASC-Quality-Reporting/; accessed October 26, 2017) has defined an acceptable accuracy threshold as 75%.

Caseload was clinically diverse, incorporating emergent, urgent, and elective procedures, inpatient and outpatient cases, and obstetric practice. All anesthesiologists were American Board of Anesthesiology Board–certified, with one exception at a secondary site (Board-eligible), no physician-anesthesiology trainees were present at any site, and more than 98% of care was delivered via an anesthesia care team model. The balance received care by physicians practicing alone.

For each procedure, available information included date and tracheal intubation outcome (successful/difficult/failed). Pediatric cases (less than 18 yr old) were excluded. Because the sample involved deidentified data, the possibility that individual patients were counted more than once could not be eliminated. In 2009, revisions to the quality assurance form added new variables including a number of patient characteristics and adverse outcomes, including some possibly related to airway management (i.e., dental injury, aspiration).
The primary events were “difficult tracheal intubation” and “failed tracheal intubation.” Definitions for these events were the same throughout the study period and consistent with other published criteria.1 “Difficult” tracheal intubation was defined as requiring at least three attempts by an experienced practitioner, and “failed” intubation was defined as requiring either surgical or percutaneous tracheotomy, cricothyrotomy, or wake-up of the patient. These definitions are as defined in the Q-CNS system and were consistent throughout the study period at all locations. For the study purposes of reporting and analysis, events classified as failed were not also included as difficult.

Statistical Analysis
The primary analysis compared rates of difficult and failed tracheal intubations at the primary facility (i.e., the large, urban community hospital). The cases were divided by procedure date into the early and late groups (early: pre-January

---

Fig. 1. Quantum Clinical Navigation System forms. (A) Handwritten report form. (B) Electronic medical record report form.
1, 2009; late: January 1, 2009 to September 30, 2015). This largest hospital was a priori selected as the source for the primary patient sample (fig. 2) due to its size and continuous use of the quality assurance system before and throughout the study period. Sample size estimates at the primary site indicated that the number of procedures available would comfortably achieve the 98,793 cases required to ensure 80% power identifying an odds ratio of 1.5 between the early and late periods. With one exception, at the other practice sites, the Q-CNS system was introduced at the start of the study period or soon thereafter. These sites were used to develop a second patient sample for similar comparison (fig. 2). To assess for different event rates between the early and late periods, comparisons employed the chi-square test, odds ratios, and 95% CI. A similar approach was applied to developing and analyzing data combined from the 15 smaller clinical sites, for comparison with findings from the primary site.

In addition to comparing the early and late periods, we were also interested in analyzing trends and annual percentage of change in outcome rates during the study period. Based on the exploratory analysis of events across years, it was inappropriate to use the ordinary least square method to determine the best-fit line. Join-point trend analysis was therefore performed to find the best-fit multisegmented line through many years of data. Join-point regression model uses the Monte Carlo permutation method to identify significant changes in the direction and magnitude over time.25 The model assumes Poisson variance and uncorrelated errors.26 We fitted a join-point regression model using difficult and failed tracheal intubation data from the primary study site, a priori combined due to the rarity of the latter, starting with no join point and evaluated whether one or more join points were statistically significant and needed to be entered into the model to best fit the data over the period of study (a maximum of two join points is permitted by default with 14 annual data points). In the final model, annual percentage change and corresponding 95% CI were estimated for each trend segment. The slopes of each neighboring pair of trend segments were also tested for statistically significant difference.

In 2009, an expanded version of the Q-CNS program was implemented at the primary site that included patient comorbidity and case information as well as limited clinical outcomes. These data were available from January 1, 2009 to December 31, 2014 and used to assess for changes in patient characteristics and potential complications of tracheal intubation such as dental injury and gastric aspiration. Descriptive summary statistics are computed as means ± SD or group frequencies (percentages) as appropriate. Chi-square tests were used to compare rates of dental injury and gastric aspiration across years (2009 to 2014). Statistical significance was set at \( P < 0.05 \). Data preparation and analyses were performed using SAS software (version 9.4; SAS Institute, Inc., USA).

Results

At the largest hospital including 421,581 eligible procedures and 1,670 difficult or failed tracheal intubation events during a 14-yr period (fig. 2; table 1), descriptive data for the patient sample is limited to the latter part of the study (2009 to 2014; table 2). Characteristics such as patient sex and age and procedure urgency during this 6-yr period remained approximately stable by year. However, relative to 2009, in 2014 more patients were American Society of Anesthesiology physical status classes III and IV (III: 51% vs 65%) and fewer I were II (50% vs 35%), suggesting an overall shift toward a sicker patient sample. In addition greater rates of

Table 1. Difficult and Failed Tracheal Intubation Events by Year and Location

<table>
<thead>
<tr>
<th>Year</th>
<th>Primary Site</th>
<th>Affiliated Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>2002</td>
<td>184 (7.5)</td>
<td>21 (1.8)</td>
</tr>
<tr>
<td>2003</td>
<td>168 (6.5)</td>
<td>67 (4.3)</td>
</tr>
<tr>
<td>2004</td>
<td>225 (8.5)</td>
<td>136 (6.8)</td>
</tr>
<tr>
<td>2005</td>
<td>197 (7.3)</td>
<td>158 (7.3)</td>
</tr>
<tr>
<td>2006</td>
<td>197 (7.8)</td>
<td>136 (6.2)</td>
</tr>
<tr>
<td>2007</td>
<td>171 (6.3)</td>
<td>105 (3.8)</td>
</tr>
<tr>
<td>2008</td>
<td>129 (4.4)</td>
<td>97 (3.1)</td>
</tr>
<tr>
<td>2009</td>
<td>113 (3.7)</td>
<td>111 (3.4)</td>
</tr>
<tr>
<td>2010</td>
<td>72 (2.2)</td>
<td>60 (1.6)</td>
</tr>
<tr>
<td>2011</td>
<td>52 (1.6)</td>
<td>55 (1.3)</td>
</tr>
<tr>
<td>2012</td>
<td>60 (1.7)</td>
<td>65 (1.5)</td>
</tr>
<tr>
<td>2013</td>
<td>41 (1.1)</td>
<td>48 (1.0)</td>
</tr>
<tr>
<td>2014</td>
<td>37 (1.0)</td>
<td>41 (0.9)</td>
</tr>
<tr>
<td>2015</td>
<td>24 (0.8)</td>
<td>32 (0.8)</td>
</tr>
<tr>
<td>Total</td>
<td>1,670 (4.0)</td>
<td>1,130 (2.6)</td>
</tr>
</tbody>
</table>

The primary study site was a single urban community hospital, and the affiliated sites were 15 small hospitals and outpatient facilities. The values represent the number of events (n) and rate per thousand cases (×1,000) at the primary (421,581 total cases) and secondary (442,428 total cases) study sites.

---

Fig. 2. Flow diagram describing procedure inclusion and exclusion for years 2002 to 2015 in developing the primary and secondary study samples.
obesity and hypertension were noted in the later years. Overall, rates of difficult and failed intubations (3.3 of 1,000 and 0.1 of 1,000, respectively) were similar to those reported in other large studies.27

In the primary analysis, a significant approximately fourfold decline in event rates over the study period was noted in comparison of both types of tracheal intubation events in the early versus late periods (difficult intubation attempts: 6.6 of 1,000 vs. 1.6 of 1,000, \(P < 0.0001\); failed intubation attempts: 0.2 of 1,000 vs. 0.06 of 1,000, \(P < 0.0001\); table 3). Equivalent early versus late analyses of data combined from the 15 smaller affiliated practices, involving 442,428 procedures (fig. 2; table 1), also demonstrated highly significant declines (difficult intubation attempts: 4.5 vs. 1.3 of 1,000, \(P < 0.0001\); failed intubation attempts: 0.3 of 1,000 vs. 0.08 of 1,000, \(P < 0.0001\); table 4). Graphic representation of annual difficult and failed tracheal intubation at the various affiliated secondary practices suggests that the general downward trend is consistent across numerous sites (fig. 3B). Data from the primary site, limited to the latter part of the study period (2009 to 2014; table 2), demonstrate no significant change in the occurrence of dental injury (\(P = 0.21\)) or gastric aspiration (\(P = 0.87\)). Patient and procedural characteristics, dental injuries, and gastric aspiration events for the 15 combined smaller affiliated sites (not shown) demonstrated findings similar to those at the primary site.

**Discussion**

In this retrospective review of a large perioperative quality assurance database over a 14-yr period ending in 2015, we observed...
steady declines in the reported rates of difficult and failed intubation at a large, urban, community hospital in the hands of experienced anesthesiologists. These findings, reflecting an approximately fourfold reduction between the first and second half of the study, were also evident at a network of 15 smaller hospitals and outpatient surgery centers. Join-point analysis revealed trend patterns that identified an approximately stable period before 2006, with the largest annual reductions between 2006 and 2010 and more gradual declines after 2010. Data available for the latter part of the study period (2009 to 2014) suggest that with each successive year, patient samples were generally sicker and more obese (i.e., possibly higher tracheal intubation risk).

Direct comparison with previous literature is challenging because there are no similar cohort studies that have examined temporal trends in difficult and failed intubation event rates. However, it is interesting to view our findings alongside published American Society of Anesthesiologists closed claim analyses, acknowledging the major differences in study design between our study and these publications. Cheney et al.,22 in a review of claims from between 1975 and 2000, observed reductions in critical respiratory events beginning in the late 1980s (23% difficult intubation). These authors speculated that the decline may have been related to the introduction of oximetry and capnography into standard anesthetic practice. A second analysis spanning 1985 to 1999 by Peterson et al.23 specific to difficult tracheal intubation noted a decline in events related to the initiation of anesthesia that occurred after 1992 (n = 179; 62 vs. 35%, P<0.05; odds ratio, 0.26; 95% CI, 0.11 to 0.63; P = 0.003) that was approximately coincident with the introduction of laryngeal mask airway devices, increased availability of fiberoptic bronchoscopy, and publication of the first American Society of Anesthesiologists expert consensus difficult tracheal intubation guidelines.7 Although these authors acknowledge the limitations of their studies, including the lack of group denominators and inability to causally link practice change with retrospective findings, they point out that successful innovations in tracheal intubation management rarely leads to an increase in claims. Unfortunately, there are no similar recent investigations to contrast with the current study.

Our data set is valuable in that it provides comprehensive peroperative data that facilitates longitudinal analysis of specific outcomes from a quality improvement perspective. However, as with the previous closed claims studies, an important limitation of our study is its retrospective design and particularly the risk for incomplete reporting that could introduce bias, potentially creating an erroneous impression of trending declines in tracheal intubation event rates (e.g., due to changes in the ease of reporting, organizational culture, fear of public transparency, etc.). Supportive evidence for the accuracy of our difficult airway data comes from a study by Walker et al.28 in which they reported that anesthesia providers may be more likely to record experiences related to airway difficulty than other events. These authors studied the validity of a voluntary anesthesia reporting process at three large clinical sites and found an overall error rate of 0.3% in reviewing 200 sequential charts for 42 items. Importantly, the error rates for administrative and demographic variables were much higher than those for quality indicators (3.0 and 1.7%, respectively, vs. 0.1%). Notably, there were zero errors related to airway management (dental injury, difficult and failed tracheal intubation, reintubation).

Other factors supporting the reliability of our findings at the primary study site include: (1) the stability of intubation event definitions, (2) the stability of documentation, data collection, and validation methodologies, (3) the quality control monitoring by dedicated staff, and (4) the corroborative evidence from secondary sites involving different providers and diverse procedural and practice characteristics. Some demographic characteristics appear to have changed during the late study period as patients became generally sicker and more often obese. Of the characteristics available, only obesity is associated with risk of difficult or failed intubation (increased), and the impact of this condition would be expected to run counter to the observed decline in intubation events. It is possible that increased obesity rates might affect adverse events related to intubation by prompting a change in practice (e.g., increased use of awake fiberoptic intubation techniques). While the perceptions of clinicians from these practices do not support this possibility, it cannot be ruled out from the available data.

Given the retrospective nature of our study, it would be inappropriate to attribute declines in difficult and failed tracheal intubation rates to specific causes. Decreasing difficult and failed intubation rates at the primary study site did overlap with a variety of technical and nontechnical perioperative innovations during the 14-yr period. Two updates

### Table 4. Occurrence of Difficult and Failed Tracheal Intubation between Early versus Late Periods at the 15 Secondary Affiliated Study Sites

<table>
<thead>
<tr>
<th>Study Period</th>
<th>n (×1,000)</th>
<th>OR (95% CI)*</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficult intubation</td>
<td>n (×1,000)</td>
<td>OR (95% CI)*</td>
<td>P Value</td>
</tr>
<tr>
<td>Early (2002–2008)</td>
<td>679 (4.5)</td>
<td>3.44 (3.04–3.90)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Late (2009–2015)</td>
<td>388 (1.3)</td>
<td>3.19 (1.92–5.30)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Failed intubation</td>
<td>39 (0.3)</td>
<td>3.19 (1.92–5.30)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Combined</td>
<td>718 (4.8)</td>
<td>3.43 (3.04–3.87)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Total cases</td>
<td>149,448</td>
<td>292,980</td>
<td></td>
</tr>
</tbody>
</table>

The values represent the number of documented events (n) and rate per thousand cases (×1,000). The *P* values were determined from chi-square test.

*OR = univariable odds ratio.*
to American Society of Anesthesiologists expert consensus algorithms were published (2003, 2013), changes in the availability of airway devices occurred at the primary site (Bullard videolaryngoscopes [Circon ACMI, USA] were available until 2010, and Glidescope videolaryngoscopes [Verathon, USA] were introduced in 2009), and other less specific activities such as simulation education, tracheal intubation training courses, changes to the process of board certification, and an increased general focus on patient safety culture also occurred during the study period. Broad supportive evidence for the importance of recent technologic innovations in airway management comes from randomized trials and meta-analyses confirming that many advanced tracheal intubation tools are superior to traditional laryngoscopy in the setting of difficult laryngoscopy, especially in the hands of trainees or novices. However, the importance of nontechnical contributions to improved airway management is also supported by studies such as the one by Berkow...
et al., which noted sustained outcome improvements after a system redesign effort to standardize the management of airway emergencies.

In summary, we present retrospective evidence of declines in the annual incidence of perioperative difficult and failed intubation events at a large community hospital and 15 affiliated practices over a 14-yr period ending in 2015, in the hands of experienced anesthesiologists. Although standard perioperative management also changed in numerous ways during the period, causal linkage of such changes to a decline in reported difficult and failed intubation events is not possible given the study design. However, many factors changed during the study period, including availability of advanced airway devices, updated expert consensus difficult intubation guidelines, training in advanced airway and crisis management, and innovations in education and certification. In addition, improvements in patient preoperative optimization and team training in general patient safety also have the potential to improve management of the difficult intubation. Such declines in difficult and failed tracheal intubation events need to be validated through assessment of other similar databases. Furthermore, better understanding of the relationship of the abovementioned factors with these observed declines may support continuing efforts to further improve intubation-related safety in settings including but also beyond the perioperative period.

Research Support
Supported by the Department of Anesthesiology of Duke University Medical Center, Durham, North Carolina.

Competing Interests
The authors declare no competing interests.

Correspondence
Address correspondence to Dr. Schroeder: Duke University Medical Center, Durham, North Carolina 27710. rebecca.schroeder@duke.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

References


