

The Journal of Education in Perioperative Medicine

ORIGINAL RESEARCH

Reporting Practices and Methodological Challenges in Anesthesiology Program Director Surveys: A Scoping Review

HAILEY JENKINS, BSPH
ADRIANA GAONA, BS

SUSAN M. MARTINELLI, MD, FASA
FEI CHEN, PHD

INTRODUCTION

Survey research is fundamental to advancing graduate medical education, yet low response rates threaten the validity and generalizability of findings. Program directors (PDs) are key stakeholders in graduate medical education, leading program administration, curriculum development, and educational quality assurance. For educators and researchers, understanding what motivates PDs to participate in surveys is essential for the development of evidence-based approaches to training, curriculum design, and educational innovation. Low response rates limit statistical power and may introduce selection bias that undermines findings used to guide policy and accreditation.

Despite the widespread use of surveys in medical education research, strategies to optimize survey methods for specific groups remain relatively unexplored. Previous research has investigated factors that enhance response rates among health care professionals in general, identifying strategies such as multiple contact attempts, monetary incentives, and mode of data collection.¹ However, these broad recommendations may not translate well to specialized populations such as residency and fellowship PDs, who juggle time constraints, survey fatigue, and competing clinical and administrative demands.^{2,3} Studies have documented low or declining survey response rates among

health care providers with physicians being particularly affected.^{1,4} Whereas some work has examined survey participation among family medicine PDs,⁴ to our knowledge, no reviews have systematically examined survey methods used in studies targeting anesthesiology PDs, including platform, length, topic relevance, endorsements, or incentives.

Researchers in anesthesiology education have surveyed PDs on a variety of topics, but the factors that drive participation remain poorly understood. Researchers often document insights about successful and unsuccessful strategies in their manuscripts, yet these lessons remain scattered throughout the literature. Consequently, fundamental decisions about survey design continue to rely on researcher preference and resources rather than empirical evidence. This gap between the critical need for PD perspectives, and the limited evidence base for effective data collection methods represents a substantial barrier to advancing educational research in anesthesiology residency and fellowship programs.

In addition to uncertainty about which design features may influence participation, variability in survey development, reporting practices, and methodological transparency may further limit the ability of researchers to learn from prior work. Differences in how surveys are constructed, described, and evaluated can obscure meaningful

interpretation of response rate patterns and hinder cumulative progress in anesthesiology education research. As such, examining methodological and reporting characteristics alongside response rate outcomes may offer insight into broader challenges facing survey-based research within graduate medical education.

This scoping review aims to (1) describe response rate patterns in surveys targeting anesthesiology PDs, (2) examine methodological characteristics that may shape participation and interpretation of findings, and (3) identify opportunities to strengthen transparency and reporting practices in survey-based anesthesiology education research.

METHODS

This scoping review incorporates the Levac enhanced 6-step methodological⁵ and PRISMA-ScR reporting standards.⁶ We selected a scoping review methodology because our aim was to map evidence across diverse study designs and identify knowledge gaps in survey methodology rather than synthesize effectiveness data, which would require a systematic review approach.

Literature Search and Study Selection

Questions were framed using the population-concept-context framework: population (anesthesiology residency and fellowship PDs, associate and assistant

continued on next page

continued from previous page

PDs), concept (survey response rates and methodological factors influencing participation), and context (medical education research in anesthesiology).⁷

A comprehensive search strategy was developed in consultation with an anesthesiology librarian at the University of North Carolina at Chapel Hill. The following databases were searched: PubMed, Embase, Scopus, ERIC, Education Full Text, CINAHL, and Proquest Health and Medical Collection. The search strategy combined terms related to anesthesiology, PDs, surveys, and response rates. The literature search was conducted from March to May 2025. The search yielded 436 unique citations after removing duplicates.

Inclusion criteria encompassed survey studies targeting anesthesiology PDs and assistant/associate PDs with national US representation, published research reporting explicit response rates, English language publications, and full-text availability. We excluded surveys targeting other populations, studies lacking response rate documentation, open-ended interviews without structured questionnaires, conference abstracts with limited detail, studies combining anesthesiology directors with other specialties inseparably, and regional or non-US programs only.

Two reviewers (HJ and AG) independently screened all 436 titles and abstracts using Covidence.⁸ Disagreements were resolved through discussion. Full-text screening was conducted by a single reviewer (HJ) with reasons for exclusion documented. A PRISMA flow diagram illustrates the selection process (Figure 1).⁹

Data Extraction and Analysis

A data extraction spreadsheet was developed based on PRISMA-ScR recommendations for scoping reviews⁶ and established survey methodology literature.^{2,3} For each included study, we recorded study characteristics (authors, year, journal, country, survey purpose, target population, study design, sample size), survey methodology details (development methods, implementation strategies, design elements), and response

rate data. We also documented whether studies reported validation processes, bias considerations, and nonresponder analyses. Factors explicitly mentioned by authors as influencing response rates were extracted and categorized. Two reviewers independently extracted data from all included studies with a mean agreement of 94% across all extracted data elements. Disagreements were resolved through discussion and reference to source materials. Data extraction was performed using Microsoft Excel.

Analysis integrated both quantitative and thematic approaches. Descriptive statistics characterized response rate distribution and methodological patterns. Response rates were analyzed according to key associations including survey platform, topic area, professional endorsement, reminder strategies, and survey length. A 12-point quality assessment framework based on established survey methodology literature^{2,3} was applied by 2 independent reviewers to evaluate survey development process, design elements, implementation strategies, and bias consideration. Studies received 0-1 point per criterion (with 0.5 for partial credit when applicable) with total scores categorized as *high* (>9 points), *moderate* (6–8.5 points), *low* (3–5.5 points), or *very low* (<3 points) quality. When survey instruments were not provided or described in the text, quality criteria requiring direct survey review were scored as 0 with this limitation noted. Full scoring criteria and further details are provided in Supplemental Item 1.

This scoping review analyzed only published literature and did not involve human subjects research; therefore, institutional review board approval was not required.

RESULTS

Study Characteristics

Our scoping review identified 36 studies meeting inclusion criteria. Twenty-one studies (58.3%) focused on anesthesiology residency programs, whereas 15 (41.7%) targeted fellowship programs: pediatric anesthesiology (5 studies, 13.9%), liver transplant anesthesiology (3 studies, 8.3%), obstetric anesthesiology (3 studies, 8.3%), pain medicine (2 studies, 5.6%), regional

and acute pain (1 study, 2.8%) and mixed fellowships (1 study, 2.8%).

Publication dates ranged from 2015 to 2025 with the highest concentration published between 2019 and 2021 (16 studies, 44.0%). Detailed characteristics of all 36 included studies are provided in Supplemental Table 2.

Response Rate Patterns

Overall response rates ranged from 19% to 82% (mean 48.3%, SD 15.7%, median 49.0%), demonstrating considerable variability across studies. Response rates fluctuated across the study period without a clear temporal trend. The highest reported response rate (82%) occurred in a 2015 survey examining pediatric anesthesia fellowship selection, whereas the lowest (19%) occurred in a 2022 survey focused on artificial intelligence.

Fellowship programs generally achieved higher mean response rates than residency programs. Among fellowships, regional and acute pain fellowship had the highest rate (71.0%, $n = 1$), followed by pediatric anesthesiology fellowships (mean 62.4%, SD 15.3%), obstetric anesthesiology (mean 56.7%, SD 4.9%), liver transplant anesthesiology (mean 54.7%, SD 7.5%), pain medicine (mean 50.4%, SD 27.8%), and mixed fellowships (31.0%, $n = 1$). Residency program surveys showed lower rates (mean 42.4%, SD 14.0%). Overall, fellowship programs achieved a significantly higher mean response rate (56.6%) than residency programs (42.4%, $p = .006$).

Methodological Factors and Response Rates

Survey platforms varied across studies. Qualtrics was the most commonly used platform ($n = 14$), followed by REDCap ($n = 9$) and SurveyMonkey ($n = 5$), whereas several other platforms were used in only 1 or 2 studies (Table 1). Four studies did not report the survey platform used.

Mean response rates varied across platforms; however, variability within categories was substantial, and several platforms were represented by only 1 or 2 studies. Qualtrics, despite being the most commonly used platform ($n = 14$), demonstrated a mean response rate of 44.3%.

continued on next page

continued from previous page

Studies with documented professional organization endorsement ($n = 12$) achieved a mean response rate of 53.7% versus 45.6% for studies without clear endorsement ($n = 24$), an absolute difference of 8.1% (Table 2). The difference was not statistically significant ($p = .137$). Endorsement occurred through professional society listservs (11 studies) and/or newsletter announcements (3 studies).

Response rates varied by survey topic (Table 3). Simulation surveys achieved the highest mean response rate (57.6%).

The relationship between email reminders and response rates showed a complex pattern (Table 4). The highest mean response rate was observed among studies reporting 3 reminders (60.3%, $n = 3$), whereas those studies reporting 1 reminder had lower rates (45.6%, $n = 5$) and those reporting 4 or more reminders averaged 48.9% ($n = 6$).

Survey length was reported by 32 studies (88.9%). Among these, surveys with 31–40 questions achieved the highest mean response rate (58.7%, $n = 4$), followed by 21–30 questions (51.4%, $n = 6$), 41–50 questions (50.6%, $n = 6$), 11–20 questions (49.5%, $n = 13$), and 0–10 questions (35.1%, $n = 3$). Studies not reporting survey length ($n = 4$) had a mean response rate of 35.7%.

Four studies (11.1%) offered lottery-based financial incentives, offering participants the chance to win gift cards valued between \$50 and \$100. These studies achieved a mean response rate of 49.0% compared with 48.2% for studies that did not report any financial incentives ($n = 32$). This 0.8% difference in response rates was not statistically significant ($p = .894$).

Survey Development, Author-Reported Considerations, and Quality

Multiple development methods were employed across studies with expert review being the most commonly reported development method (18 studies, 50.0%), followed by pilot testing (14 studies, 38.9%) and literature review (7 studies, 19.4%). Two studies (5.6%) reported using focus groups, and 3 studies (8.3%) reported individual interviews with

stakeholders during survey development. Two studies (5.6%) provided no development information. Twenty-seven studies (75.0%) created new instruments, whereas 6 studies adapted existing tools with published validity evidence (16.7%).

Surveys typically employed multiple question formats. Multiple choice was the most common question type (30 studies, 83.3%), followed by open-ended/free-text questions (28 studies, 77.8%), demographic questions (25 studies, 69.4%), yes/no questions (23 studies, 63.9%), Likert rating scales (20 studies, 55.6%), ranking questions (6 studies, 16.7%), and numerical rating scales (4 studies, 11.1%). Most surveys (33 studies, 91.7%) combined multiple question formats within a single instrument, whereas 1 study used only multiple-choice questions. Two studies (5.6%) did not specify their question types or provide their survey for review.

Most studies (23 studies, 63.9%) acknowledged suboptimal response rate as a study limitation. Authors also cited methodological limitations as potentially impacting response rate, including lack of instrument validation (10 studies, 27.8%) and absence of pilot testing (4 studies, 11.1%). Regarding barriers to participation, authors most often theorized that timing issues with survey distribution (5 studies, 13.9%) and survey fatigue (4 studies, 11.1%) may have impacted response rates. Additional barriers mentioned include pandemic disruptions (2 studies, 5.6%) and time constraints experienced by PDs (2 studies, 5.6%). In contrast, few authors explicitly identified factors potentially improving their response rates with only survey brevity (2 studies, 5.6%) and multiple email reminders (2 studies, 5.6%) mentioned more than once. Most studies neither discussed factors they believed enhanced participation nor provided recommendations for future survey research.

Systematic quality assessment using our 12-point framework revealed varying quality ratings across included studies. Five studies (13.9%) met high-quality criteria (≥ 9 points), 19 studies (52.8%) demonstrated moderate quality (6–8.5 points), 10 studies (27.8%) were classified

as low quality (3–5.5 points), and 2 studies (5.6%) fell into the very low-quality category (< 3 points). Assessment of question quality, structure, and response options was limited as 15 studies (41.7%) did not provide their full survey instrument in the manuscript or appendices though 4 of these described survey questions within the manuscript text.

DISCUSSION

This scoping review identified substantial variability in response rates among surveys of anesthesiology PDs with fellowship programs achieving significantly higher participation than residency programs. The overall response rate of 48.3% aligns with health care professional survey literature¹⁰ though it falls below the 70% threshold at which nonresponse bias becomes problematic.¹¹ The higher response rates among fellowship programs (56.6%) compared with residency programs (42.4%) may reflect structural differences between these populations. Fellowship communities often involve fewer trainees and more tightly connected professional networks, which may increase perceived relevance or collegial obligation to participate in surveys. Fellowship program directors may be more likely to respond to surveys from colleagues with whom they have direct professional relationships.

We did not observe a statistically significant association between professional organization endorsement and response rate or between the provision of incentives and response rate. Previous studies have found that organizational sponsorship improves participation.¹² Although professional organization endorsement showed a descriptive 8.1% advantage in our review, this difference was not statistically significant ($p = .137$). Despite recommendations for financial incentives,^{12,13} only 4 studies offered compensation, possibly reflecting budget constraints or uncertainty about effectiveness in this population. The modest incentive amounts may or may not have reached the effectiveness threshold for this population, representing a key area for future research to determine optimal

continued on next page

continued from previous page

incentive levels and whether guaranteed or lottery-based compensation is more effective.

Although we also descriptively summarized additional design considerations, such as survey platform and topic area, we did not conduct hypothesis testing for these factors because some categories included small numbers of studies. Consequently, it remains unclear whether the notable absolute differences observed across certain categories are statistically significant. The relationship between reminder frequency and response rate requires cautious interpretation. At first glance, the finding that 3 reminders appeared “optimal” whereas 4 or more reminders were associated with lower response rates may seem counterintuitive. However, several methodological limitations temper this interpretation. First, these findings are observational and correlational. The number of reminders was not randomly assigned but likely determined in response to interim participation rates. In practice, studies with poor initial response may have attempted more reminders as a salvage strategy, confounding the interpretation. The observed association likely reflects reactive survey administration decisions rather than a true decrease in participation caused by additional reminders. Without data on response rates after each successive reminder, we cannot determine the incremental value of additional reminders.

Survey length also demonstrated a counterintuitive pattern with surveys containing 31–40 questions achieving higher response rates than shorter instruments. Although shorter surveys are often assumed to improve participation, longer surveys may signal greater perceived importance or institutional relevance, particularly when endorsed by professional organizations or addressing high-priority educational topics. Alternatively, very brief surveys may be perceived as less impactful or less relevant to program leadership decisions. These explanations remain speculative and may reflect unmeasured factors such as timing of distribution, competing survey burden, topic salience, or existing

professional relationships rather than survey length itself. Survey design features such as length are also unlikely to operate independently, and observed patterns may reflect interactions among multiple factors rather than the isolated effect of any single variable. Future research using experimental study designs is warranted to examine these factors collectively. Additionally, survey length may not be directly tied to the time commitment for the survey. A short survey with multiple parts to each question or open answered questions may take longer to complete than a longer survey with quick multiple-choice questions.

Descriptive variation in response rates was observed across platforms though hypothesis testing was not conducted due to sample size. Whereas previous research has examined survey delivery modes (postal versus online),^{13,14} our findings highlight an unexplored area: direct comparisons between online platforms themselves, a gap that becomes increasingly relevant as digital surveys dominate medical education research. To our knowledge, there are no prior studies directly comparing survey platform performance among anesthesiology PDs or similar graduate medical education leadership populations. As such, differences observed in this review should be interpreted cautiously and viewed within the broader context of variable reporting practices rather than as evidence of platform effectiveness though this represents a potential area for future investigation.

Survey topic also emerged as a factor of interest with notable descriptive differences in response rates. Surveys on simulation showed the highest response rates (57.6%), whereas a single survey on artificial intelligence showed the lowest (19.6%). For clarity, surveys classified as “simulation” represented multiple studies addressing simulation in anesthesiology training. Because simulation-based education encompasses diverse instructional and curricular applications, this category should be interpreted as heterogeneous rather than a single unified topic area. Whereas hypothesis testing was not conducted, the magnitude of this difference may

carry educational significance. These differences also raise questions about whether certain topics align with PD roles as some survey topics may be better suited for other departmental experts rather than PDs themselves, raising questions about whether respondents were encouraged to delegate surveys requiring specialized knowledge and whether this affected dismissal rates. Although surveys directed at PDs are typically intended to be relevant to their leadership roles; perceived actionability, administrative burden, and alignment with program priorities may influence whether directors choose to engage with specific topics. PDs may also be less inclined to complete surveys when the anticipated knowledge gains seem irrelevant to their program needs.

The limited author discussion of success factors reveals a critical blind spot in survey methodology research. Whereas authors frequently acknowledged suboptimal response rate as a limitation (63.9% of studies), few attempted to explain what drove or hindered participation rates. This disconnect between widespread acknowledgment of limitations and sparse discussion of solutions suggests researchers may not understand what influences PD participation. The barriers authors did identify, such as time constraints or survey fatigue, represent researcher interpretations rather than direct feedback from nonrespondents. Without input from PDs who chose not to participate or methodologically rigorous studies examining response patterns in this population, researchers continue operating on untested assumptions, potentially perpetuating ineffective strategies.

Survey reporting quality presents additional concerns. Whereas most studies acknowledged some methodological limitations, few provided the transparency needed for collective learning. The absence of accessible survey instruments in 41.7% of included studies prevents assessment of survey quality, a factor that may significantly influence response and completion rates and hinders research replication. This lack of transparency aligns with previous findings documenting inconsistent reporting standards in

continued on next page

continued from previous page

survey research,¹⁵ particularly within anesthesia journals.¹⁶ This creates an unfortunate cycle: researchers cannot learn from others' successes and failures, leading to repeated mistakes and missed opportunities for methodological advancement. Future studies should consider staged response rate reporting to distinguish between effective strategies and selection bias, enabling evidence-based recommendations for surveying this population. Additionally, improved adherence to existing reporting standards for survey research such as CROSS¹⁷ would enhance transparency in instrument development processes and enable meaningful comparisons across studies. Based on the gaps identified in this review, we propose a set of reporting considerations for survey-based medical education research (Box 1).

Our conservative scoring approach may have underestimated actual survey quality as only explicitly reported methods received credit. Only 19.4% of studies reported conducting a literature review; whether additional preparatory work occurred but went unreported cannot be determined from published manuscripts. This gap between actual practice and reported methods further underscores the need for adoption of standardized reporting requirements that capture the full scope of survey development procedures.

Our study has limitations. Relying on author-reported factors—such as speculations about time constraints or survey fatigue offered in individual study discussions—rather than direct input from PDs may miss key factors in influencing participation. Our relatively small sample ($n = 36$) spans 10 years during which survey practices may have evolved, and the descriptive nature of this scoping review precludes inferential analysis. As a result, observed differences may reflect variability across studies rather than meaningful effects of individual survey design elements. Inconsistent reporting of survey methodology prevented systematic quality assessment. Publication bias may favor studies with higher response rates. Additional contextual variables not

consistently captured in published reports (including time of year, survey language and formatting, font size, investigator reputation, concurrent survey burden, and ease of survey accessibility) may also influence participation but could not be evaluated within this review. Researcher reputation, institutional relationships, and survey timing relative to competing requests may ultimately exert greater influence than isolated design choices.^{3,18,19}

Rather than establishing definitive predictors of response behavior, our findings highlight broader challenges in survey design transparency, reporting practices, and interpretation across anesthesiology education research. This review, therefore, focuses on describing patterns and identifying methodological gaps across studies rather than inferring causal relationships between specific survey features and participation outcomes. Framing these patterns as lessons, particularly regarding shortcomings in reporting survey procedures and contextual factors, may, therefore, provide greater long-term value to future researchers than isolated comparisons of response rates alone. By addressing these reporting and design limitations, future researchers may be better positioned to build cumulative, generalizable knowledge rather than repeatedly confronting the same limitations.

Overall, this review suggests that a key opportunity to improve survey-based research in anesthesiology education lies not in identifying a single optimal tactic but in strengthening transparency, reporting consistency, and methodological rigor across studies. By framing observed response rate patterns as lessons about survey design and reporting shortcomings, future investigators may be better positioned to develop evidence-informed approaches that reflect the realities facing busy clinician educators. Future research should directly engage PDs through interviews, focus groups, and experimental testing to understand their perspectives on survey burden, timing, platforms, and motivating factors, ultimately developing evidence-based guidelines that address both researcher needs and the realities facing busy clinician educators. Our

findings also highlight critical gaps in survey reporting practices with many studies omitting survey instruments and methodological details essential for advancing the field. Journals and investigators may consider adopting standardized reporting frameworks and practical guidance, such as inclusion of survey instruments and staged response rate reporting, to enable cumulative learning and methodological advancement of survey-based medical education research.

CONCLUSION

This scoping review found substantial variation in response rates among anesthesiology PDs, with fellowship directors showing higher participation than residency directors and descriptive variability observed across survey design features. Rather than identifying definitive predictors of participation, these findings underscore the need for improved transparency and consistency in survey methodology reporting across anesthesiology education research.

References

1. Cho YI, Johnson TP, VanGeest JB. Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response. *Eval Health Prof.* 2013;36(3):382-407.
2. Sullivan GM, Artino AR. How to create a bad survey instrument. *J Grad Med Educ.* 2017;9(4):411-5.
3. Artino AR, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE guide no. 87. *Med Teach.* 2014;36(6):463-74.
4. Ericson A, Bonuck K, Green LA, et al. Optimizing survey response rates in graduate medical education research studies. *Fam Med.* 2023;55(5):304-10.
5. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci.* 2010;5(1):69.
6. Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* 2018;169(7):467-73.
7. Peters MD, Godfrey C, McInerney P, et al. Scoping reviews. In: Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z, eds. *JBI Manual for Evidence Synthesis*. JBI; 2024.
8. Covidence Systematic Review Software. Published online 2025. <https://www.covidence.org/>
9. Page MJ, McKenzie JE, Bossuyt PM, et al. The

continued on next page

continued from previous page

- PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021:n71.
10. Meyer VM, Benjamens S, Moumni ME, Lange JFM, Pol RA. Global overview of response rates in patient and health care professional surveys in surgery: a systematic review. *Ann Surg*. 2022;275(1):e75-81.
 11. Prince M. Epidemiology. In: *Core Psychiatry*. Elsevier; 2012:115-29.
 12. VanGeest JB, Johnson TP, Welch VL. Methodologies for improving response rates in surveys of physicians: a systematic review. *Eval Health Prof*. 2007;30(4):303-21.
 13. Martins Y, Lederman RI, Lowenstein CL, et al. Increasing response rates from physicians in oncology research: a structured literature review and data from a recent physician survey. *Br J Cancer*. 2012;106(6):1021-6.
 14. Harrap B, Taylor T, Russell G, Scott A. A randomised controlled trial of email versus mailed invitation letter in a national longitudinal survey of physicians. *PLoS One*. 2023;18(8):e0289628.
 15. Guidolin K, Wexner SD, Jung F, et al. Strengths and weaknesses in the methodology of survey-based research in surgery: a call for standardization. *Surgery*. 2021;170(2):493-8.
 16. Story DA, Gin V, na Ranong V, et al. Inconsistent survey reporting in anesthesia journals. *Anesth Analg*. 2011;113(3):591-5.
 17. Sharma A, Minh Duc NT, Luu Lam Thang T, et al. A consensus-based checklist for reporting of survey studies (CROSS). *J Gen Intern Med*. 2021;36(10):3179-87.
 18. Brtnikova M, Crane LA, Allison MA, et al. A method for achieving high response rates in national surveys of U.S. primary care physicians. *PLoS One*. 2018;13(8):e0202755.
 19. Morris CJ, Cantrill JA, Weiss MC. GP survey response rate: a miscellany of influencing factors. *Fam Pract*. 2001;18(4):454-6.

The authors are at the University of North Carolina in Chapel Hill, NC. **Hailey Jenkins** is a medical student in the School of Medicine, **Adriana Gaona** is an MPH student at the Gillings School of Global Public Health, and **Susan M. Martinelli** is a Professor and **Fei Chen** an Associate Professor in the Department of Anesthesiology.

Corresponding author: Hailey Jenkins, BSPH, University of North Carolina at Chapel Hill, 150 Medical Drive, Chapel Hill, NC 27514. Telephone: (219) 484-0917

Email address: Hailey Jenkins: Hailey_jenkins@med.unc.edu

Funding: The authors report no external funding source for this review.

Conflicts of interest: The authors declare that they have no competing interests.

Previous Presentations: Preliminary results from this study were presented as a poster at the University of North Carolina Academy of Educators Conference, June 2025, Chapel Hill, North Carolina.

Abstract

Background: Survey research is widely used in anesthesiology education, yet response rates among program directors (PDs) remain variable and difficult to interpret. Inconsistent reporting of survey methods may limit meaningful comparisons and hinder methodological advancement. This scoping review examined response rate patterns and reporting practices in surveys targeting anesthesiology PDs.

Methods: We conducted a scoping review to map evidence and identify gaps in survey methodology. Seven databases were searched for surveys of US anesthesiology

PDs, yielding 436 citations screened by 2 reviewers. Data extraction included study characteristics, methodology, and response rates. Quality was assessed with a 12-point framework of survey development, design, implementation, and bias consideration.

Results: Thirty-six studies were included with response rates ranging from 19% to 82% (mean 48.3%). Fellowships averaged 56.6% versus residencies 42.4%. Average rates by platform were Google Forms 64.8%, REDCap 51.0%, SurveyMonkey 51.0%, and Qualtrics 44.3%. Studies with 3 reminders averaged 60.3%, and those with 4 or more averaged 48.9%. Among topics, simulation achieved the highest average response rate (57.6%). Survey instruments were unavailable in 41.7% of studies. Quality ratings were high in 13.9%, moderate in 52.8%, and low or very low in 33.3%.

Conclusions: Through analysis of key design considerations, this review highlights variability in reporting practices and methodological transparency across anesthesiology PD surveys. Methodological features such as survey platform, reminders, and topic varied across studies, but inconsistent reporting limited interpretation of the implications of these methodological choices. Future work should prioritize consistent reporting standards and prospective controlled studies to better inform survey design in anesthesiology education.

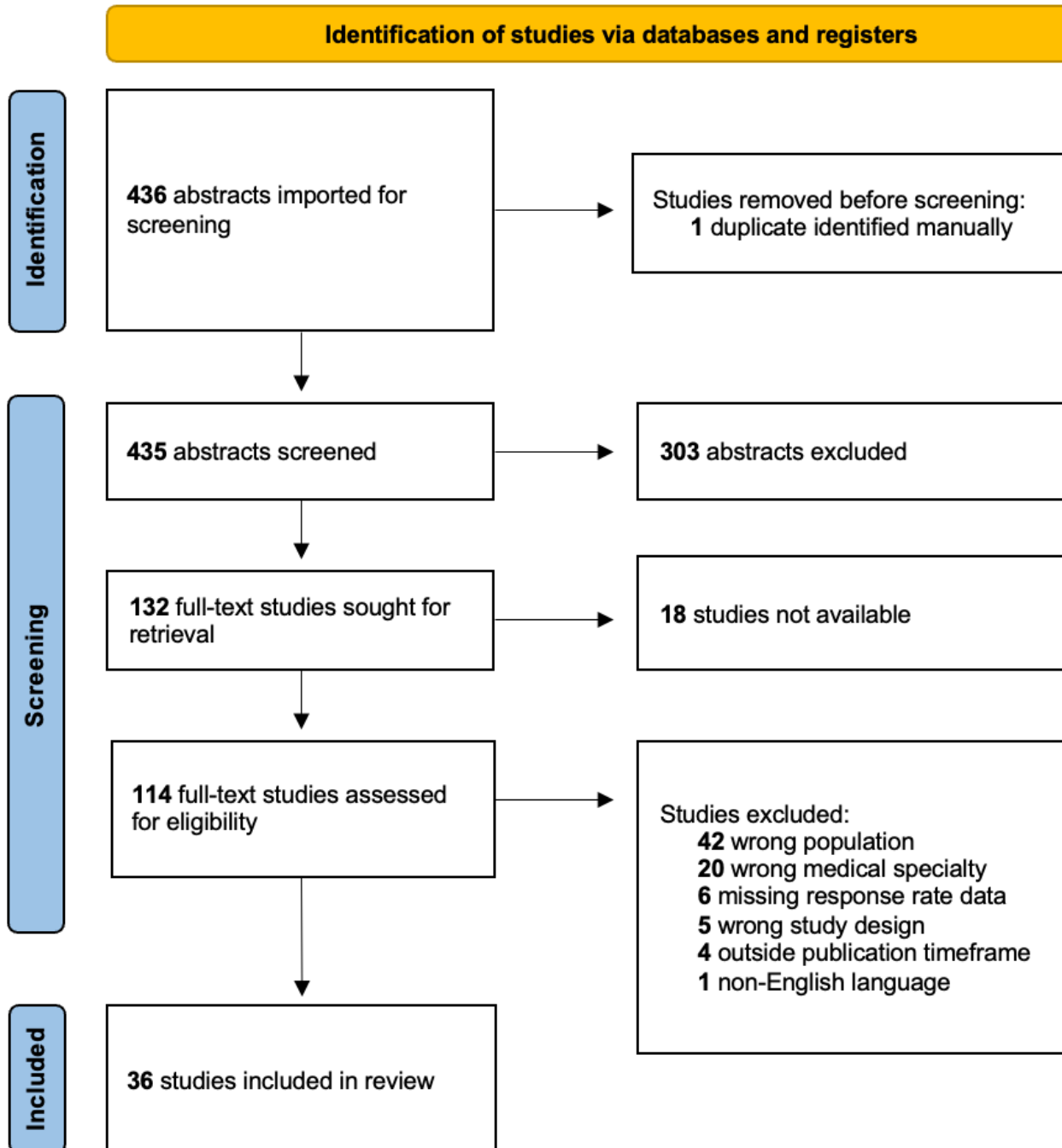
Keywords: Surveys and questionnaires, education, medical, graduate, internship and residency, anesthesiology, scoping review, program director, response rates

continued on next page

continued from previous page

Figures

Figure 1. PRISMA 2020 flow diagram showing study selection process.



continued on next page

continued from previous page

Figures

Box 1. *Reporting considerations for survey-based medical education research*

Based on the gaps identified in this scoping review, the following practices may improve transparency and interpretability of surveys involving clinical educators:

- Provide the full survey instrument or representative items to support appraisal and replication.
 - » Describe survey development processes (eg, literature review, expert review, pilot testing, or adaptation of existing tools).
 - » Clearly define the sampling frame, eligibility criteria, recruitment pathways, and response rate denominator.
 - » Report survey administration details, including platform, timing of distribution, reminder strategy, and staged response rates when available.
 - » Report whether incentives were offered, including type (financial vs nonfinancial), amount, and delivery method (guaranteed vs lottery)
 - » Address potential sources of nonresponse bias and describe any nonresponder analyses.
 - » Interpret descriptive response rate patterns cautiously when inferential analyses are not performed.

continued on next page

continued from previous page

Tables

Table 1. Response Rates by Survey Platform

Platform	Number of Studies	Mean Response Rate, %	Median Response Rate, %	SD
Google Forms	2	64.8	64.8	8.8
SelectSurveys	1	60.0	60.0	/
QuestionPro	1	56.0	56.0	/
REDCap	9	51.0	51.0	12.8
SurveyMonkey	5	51.0	44.0	23.3
Qualtrics	14	44.3	41.1	14.8
Not reported	4	39.5	40.0	17.1

Table 2. Impact of Professional Endorsement on Response Rates

Endorsement Status	Number of Studies	Mean Response Rate, %	Median Response Rate, %	SD
Professional organization endorsement	12	53.7	53.5	14.1
No endorsement or unclear	24	45.6	45.0	16.1

Table 3. Response Rates by Survey Topic

Survey Topic	Number of Studies	Mean Response Rate, %	Median Response Rate, %	SD
Simulation	4	57.6	62.5	19.3
Transplant anesthesia	2	52.5	52.5	9.2
Ultrasound/POCUS	4	51.3	57.8	13.8
Education/curriculum	9	51.2	53.0	13.7
Recruitment processes	9	46.7	48.9	20.2
DEI	4	44.5	43.5	12.5
Global health	1	44.0	44.0	/
Mentorship	2	37.6	37.6	5.0
Artificial intelligence	1	19.6	19.6	/

Table 4. Relationship Between Email Reminders and Response Rates

Reminder Number	Number of Studies	Mean Response Rate, %	Median Response Rate, %	SD
1 Reminder	5	45.6	48.0	10.8
2 Reminders	16	47.7	47.6	17.5
3 Reminders	3	60.3	57.0	13.3
4+ Reminders	6	48.9	51.0	11.5
Not reported	6	45.4	36.4	20.3

continued on next page

continued from previous page

Supplemental Online Materials

Supplemental Item 1

Survey Quality Assessment Framework

Studies were evaluated across 12 quality criteria developed using established survey methodology standards.^{2,3} Each criterion received 0-1 point, with partial credit (0.5 points) available for select criteria, maximum 12 points.

1. Literature review conducted: Yes (1 point) or No (0 points)
2. Expert review conducted: Yes (1 point) or No (0 points)
3. Pilot testing conducted: Yes (1 point) or No (0 points)
4. No leading/biased questions: None identified (1 point), 1-2 minor issues (0.5 points), multiple problems or survey unavailable (0 points)*
5. No double-barreled questions: Yes (1 point) or No (0 points)*
6. Balanced response options: Well-balanced (1 point), minor issues (0.5 points), significant imbalances or survey unavailable (0 points)*
7. Multiple contact attempts: Yes (1 point) or No (0 points)
8. Timing consideration: Mentioned consideration of timing for survey administration (e.g., avoiding holidays, recruitment seasons, or academic calendar conflicts) (1 point) or not discussed (0 points)
9. Survey length appropriate: ≤ 20 questions or ≤ 10 minutes (as reported by authors or stated to participants) (1 point), 21-35 questions or 10-15 minutes (0.5 points), > 35 questions or > 15 minutes or not reported (0 points)**
10. Response rate calculation clear: Yes (1 point) or No (0 points)
11. Non-responder analysis conducted: Yes (1 point) or No (0 points)
12. Bias/limitations discussed: Thoroughly (1 point), briefly in 1-2 sentences (0.5 points), not discussed (0 points)

continued on next page

continued from previous page

Supplemental Online Materials

*Full survey availability enables quality assessment and replication. When surveys were unavailable for review, design quality criteria (items 4-6) were scored as zero.

**When both question count and completion time reported, more conservative score assigned. For question ranges, maximum value used.

The scoring system prioritized conservative assessment, ensuring that quality scores reflected the minimum demonstrated quality rather than assumed best-case scenarios. This approach aimed to provide a fair but rigorous evaluation of survey methodology across the included studies while accounting for variations in reporting completeness.

Quality Categories:

- High: ≥ 9 points
- Moderate: 6-8.5 points
- Low: 3-5.5 points
- Very low: < 3 points

continued on next page

continued from previous page

Supplemental Online Materials

Supplemental Table 2. Characteristics and Quality Scores for All Included Studies (N=36)

First Author, Year	Survey Topic	Target Population	N Invited	N Responded	Response Rate	Survey Platform	No. Reminders	Endorsement	Incentives	Survey Length (Items)	Instrument Available	Quality Score	Quality Rating
Johnstone, 2023	Application/Selection Process	Residency	163	31	19.0%	NS*	2	No	No	1-10	Partial	5	Low
Elliott, 2024	Artificial Intelligence	Residency	163	32	19.6%	Qualtrics	2	No	No	NS*	No	4	Low
Vinagre, 2020	Application/Selection Process	Residency	159	45	28.0%	Qualtrics	4+	No	No	11-20	Yes	8.5	Moderate
Streiff, 2018	Education/Curriculum	Residency	130	39	30.0%	SurveyMonkey	1	No	No	11-20	Yes	6.5	Moderate
Chen, 2022	Simulation	Residency	148	45	30.4%	REDCap	2	No	No	1-10	Partial	8	Moderate
Conway, 2015	Ultrasound/POCUS	Fellowship	101	31	30.7%	SurveyMonkey	NS*	No	No	NS*	No	2.5	Very low
Kraus, 2022	DEI	Fellowship	327	101	31.0%	Qualtrics	NS*	Yes	No	41-50	Yes	6.5	Moderate
Pierre, 2021	Application/Selection Process	Residency	183	62	33.8%	Qualtrics	NS*	No	No	NS*	No	1.5	Very low
Gonzalez, 2016	Mentorship	Residency	131	45	34.0%	NS*	2	No	No	21-30	Yes	5.5	Low
Huang, 2020	Education/Curriculum	Residency	156	56	36.0%	Qualtrics	2	Yes	Yes - Lottery	11-20	Yes	7	Moderate
Ho, 2021	Application/Selection Process	Residency	142	46	37.0%	REDCap	2	No	No	11-20	Yes	8.5	Moderate
Sharpe, 2021	DEI	Residency	145	56	39.0%	Qualtrics	NS*	Yes	No	41-50	Partial	6.5	Moderate
Rinehart, 2017	Education/Curriculum	Residency	132	55	41.0%	Qualtrics	1	No	No	11-20	Yes	7	Moderate
Gupta, 2017	Mentorship	Residency	129	53	41.1%	Qualtrics	2	No	Yes - Non-financial	14-21	No	6.5	Moderate
Kaur, 2017	Global Health	Residency	128	56	44.0%	SurveyMonkey	2	No	No	11-20	Yes	7	Moderate
Yoon, 2021	Transplant Anesthesia	Fellowship	137	61	46.0%	NS*	4+	No	No	21-30	No	4	Low
Krueger, 2024	DEI	Fellowship	69	33	48.0%	REDCap	1	Yes	No	21-30	Yes	8	Moderate
Dutoit, 2024	Application/Selection Process	Residency	174	85	48.9%	REDCap	4+	Yes	No	31-40	Yes	9	High
Isaak, 2017	Education/Curriculum	Residency	136	67	49.0%	Qualtrics	3	No	Yes - Lottery	41-50	Yes	9	High
DiGiusto, 2021	Application/Selection Process	Fellowship	60	30	50.0%	Qualtrics	1	Yes	No	41-50	No	5.5	Low
Ruiz, 2025	Application/Selection Process	Fellowship	45	23	51.0%	REDCap	2	No	Yes - Lottery	31-40	Yes	6	Moderate
Raymond, 2019	Education/Curriculum	Residency	150	79	53.0%	REDCap	4+	No	No	11-20	Yes	7.5	Moderate
Chen, 2021	Education/Curriculum	Residency	156	87	56.0%	Question Pro	NS*	No	No	1-10	Yes	9	High
Wolpaw, 2020	Education/Curriculum	Residency	149	85	57.0%	Qualtrics	3	No	No	11-20	Yes	9	High
O'Brien, 2024	Ultrasound/POCUS	Fellowship	58	33	57.0%	REDCap	2	Yes	No	21-30	No	5.5	Low
Pustavoitau, 2023	Ultrasound/POCUS	Residency	147	86	58.5%	Google Forms	4+	No	No	NS*	No	3.5	Low
Dominguez, 2019	Ultrasound/POCUS	Fellowship	81	29	59.0%	REDCap	1	Yes	No	11-20	No	3.5	Low
Crouch, 2021	Transplant Anesthesia	Fellowship	110	65	59.0%	Qualtrics	4+	Yes	No	11-20	Partial	8	Moderate
Maffucci, 2023	Simulation	Fellowship	22	13	59.0%	NS*	2	No	No	11-20	No	3.5	Low
MacCormick, 2019	DEI	Fellowship	50	30	60.0%	SelectSurvey	2	No	Yes - Lottery	31-40	Yes	9.5	High
Rochlen, 2016	Simulation	Residency	132	87	66.0%	Qualtrics	2	No	No	41-50	Yes	6.5	Moderate
Brzezinski, 2020	Education/Curriculum	Residency	117	80	68.4%	SurveyMonkey	2	No	No	41-50	Yes	8.5	Moderate
Kohan, 2021	Education/Curriculum	Fellowship	100	70	70.0%	Qualtrics	2	Yes	No	11-20	Yes	7.5	Moderate
Lim, 2024	Application/Selection Process	Fellowship	79	56	71.0%	Google Forms	2	Yes	No	11-20	No	3.5	Low
Ambardekar, 2020	Simulation	Fellowship	60	45	75.0%	REDCap	3	Yes	No	31-40	Yes	6.5	Moderate
Benzon, 2015	Application/Selection Process	Fellowship	51	42	82.0%	SurveyMonkey	NS*	No	No	21-30	Yes	6	Moderate

*NS = Not Specified (value not reported in manuscript). Survey instrument availability was assessed at the time of data extraction using open-access sources; instruments may be available via institutional journal access or paid subscriptions.