Critical Appraisal of the Quality of Publications in Hepatobiliary and Pancreatic Surgery Research Using the American College of Surgeons NSQIP Database

Diana Hockett, MD, Judy B Rabinowitz, MLIS, Yong K Kwon, MD, FACS, Bellal Joseph, MD, FACS, Haytham Kaafarani, MD, FACS, Hassan Aziz, MD

BACKGROUND: The use of the American College of Surgeons (ACS) NSQIP has increased in hepatobiliary and pancreatic surgery (HPB) research as it provides access to high-quality surgical outcome data on a national scale. Using the ACS NSQIP database, this study examined the methodologic reporting of HPB publications.

STUDY DESIGN: Web of Science core collection (all editions) was queried for all HPB studies using the ACS NSQIP database published between 2004 and 2022. In addition, a critical appraisal was performed using the *JAMA Surgery* checklist, Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement, and Reporting of Studies Conducted Using Observational Routinely-Collected Health Data (RECORD) statement.

RESULTS: In total, 276 studies were included in the analysis. The median scores for the *JAMA Surgery* checklist, RECORD statement, and STROBE statement were 4 of 7 (interquartile range [IQR] 3 to 5), 3 of 10 (IQR 2 to 4), and 15 of 21 (IQR 13 to 17), respectively. The criteria with the highest rates of nonadherence were discussing competing risks, clear definitions of inclusion and exclusion criteria, unadjusted and adjusted outcomes, provision of supplementary data, and performing subgroup analyses. Additionally, when examining checklist fulfillment of hepatobiliary studies and pancreatic studies separately, pancreatic studies demonstrated significantly greater fulfillment of the STROBE statement checklist items.

CONCLUSIONS: Satisfactory reporting of methodology is present among HPB studies utilizing the ACS NSQIP database, with multiple opportunities for improvement. Areas for improved adherence include discussing competing risks, providing supplementary information, and performing appropriate subgroup analysis. Given the increasing role of large-scale databases in surgical research, enhanced commitment to reporting guidelines may advance HPB research and ensure quality reporting. (J Am Coll Surg 2023;236:449–460. © 2022 by the American College of Surgeons. Published by Wolters Kluwer Health, Inc. All rights reserved.)

Recently, a great emphasis has been placed on quality improvement initiatives to reduce treatment costs and improve surgical outcomes.[1] As a result of new policies requiring physicians to monitor their quality through recognized quality indicators, large databases have been developed, such as the NSQIP from the American College of Surgeons (ACS).[2,3] It represents the largest risk-adjusted, quality improvement database that surgeons prepare, edit, and authenticate. NSQIP was first conceived in 1998 and implemented in 2003.[4]

The ACS NSQIP database is frequently utilized by surgeons to report clinical outcomes.[5] Similarly, more original research articles are being published in the field of hepatobiliary and pancreatic (HPB) surgery using data from

Disclosure Information: Nothing to disclose.

Received August 5, 2022; Revised September 25, 2022; Accepted October 16, 2022.

From the Tufts University School of Medicine (Hockett), and Hirsh Health Sciences Library, Tufts University (Rabinowitz), Boston, MA; Division of Transplant and Hepatobiliary Surgery, University of Southern California, Los Angeles, CA (Kwon); Division of Trauma, College of Medicine, University of Arizona, Tucson, AZ (Joseph); Division of Trauma, Massachusetts General Hospital, Boston, MA (Kaafarani); and Division of Transplant and Hepatobiliary Surgery, University of Iowa Hospitals and Clinics, Iowa City, IA (Aziz).

Correspondence address: Hassan Aziz, MD, University of Iowa Hospital and Clinics, 200 Hawkins, C41-S GH, Iowa City, IA 52242. email: Hassan-aziz@uiowa.edu

Supplemental digital content is available for this article.

https://doi.org/10.1097/XCS.0000000000000477
ISSN 1879-1190/22

Abbreviations and Acronyms

HPB	=	hepatobiliary and pancreatic surgery
IF	=	impact factor
IQR	=	interquartile range
RECORD	=	Reporting of Studies Conducted Using Observational Routinely-Collected Health Data
STROBE	=	Strengthening the Reporting of Observational Studies in Epidemiology

the ACS NSQIP database.[6] However, the growing use of large databases in surgical research has generated several concerns about the methodologic quality of these studies and has raised questions about potential limitations of the database.[6] Hence, to ensure quality and adherence to established reporting standards, guidelines on methodologic reports have been devised.[7] Unfortunately, compliance with these reporting statements has not always been optimal.[8] Therefore, assessing the methodologic quality of studies using these large databases is imperative so that we can better understand the reasons behind suboptimal compliance, identify inadequacies, and propose novel solutions for improving reporting so that we can improve the process.[9]

The ACS NSQIP database allows researchers to conduct studies related to a vast diversity of surgical topics, including HPB.[10] Access to this large-scale database has allowed for the development of predictive outcome models and examination of nationwide surgical practice trends over time. In addition, HPB surgeons specifically have used this database to track and improve HPB outcomes. Therefore, it is vital to monitor the impact of large database usage in HPB research, ensuring high-quality research. While there has been an increase in the number of HPB studies published using the ACS NSQIP database, there is still a lack of research assessing the methodologic quality of these publications as a whole. The purpose of this study is to critique the methodologic quality of HPB publications based on the ACS NSQIP database in order to assess their scientific merit. The secondary objective of this research is to identify areas of improvement in HPB research as a means of advancing this field for the future.

METHODS

Eligibility and exclusion criteria

The ACS NSQIP database was searched in order to find all the publications involving this database starting in 2004 all the way until February 19, 2021 as the time frame from its inception to this date. For the purposes of this study, we used Clarivate's Web of Science Core Collection (all editions, including Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index, Emerging Sources Citation Index, Conference Proceedings Citation Index, Book Citation Index, Current Chemical Reactions, and Index Chemicus). A topic search for HPB procedures was conducted using the following keyword combinations: (liver* OR hepat* OR pancreas* OR gallbladder*) and (surgical* OR surger* OR transplant*) OR (hepatectomy OR pancreatectomy OR pancreaticoduodenectomy OR cholecystectomy, bile duct reconstruction). To narrow the results based on utilization of the NSQIP database, a topic search was also conducted using the following keywords: (NSQIP OR "National Surgical Quality Improvement Program"). The Boolean operator "AND" was used with each term to add the terms "ACS NSQIP," "NSQIP," "American College of Surgeons National Surgical Quality Improvement Program," and "National Surgical Quality Improvement Program." No language limitations were imposed. Studies using other databases such as the ACS NSQIP Pediatric Database, National Readmission Database, Surveillance, Epidemiology, and End Results, National Inpatient Sample, and National Cancer Database were excluded. Articles comparing outcomes from the ACS NSQIP database with institutional-based data were also excluded. Commentaries, letters to the editor, review articles, and editorials were excluded.

Study selection

The titles and abstracts of all the identified ACS NSQIP publications were independently reviewed by 2 authors (HA and DH) for the possibility of eligibility for inclusion in the program. If at least 1 of the 2 authors deemed a publication to be eligible for retrieval, then the full text of the publication was obtained. The full text of the study criteria was independently reviewed by both authors. A third independent author (YK) coordinated with other authors if there were any disagreements between the authors. The process is summarized in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram shown in Figure 1.

Data extraction and processing

The following information was obtained from each publication: author(s), title, year of publication, name of publishing journal, and HPB subspecialty.[11] In addition, impact factors (IF) (2021) of publishing journals were extracted from the Web of Science's Journal Citation Reports. Journal policies regarding author requirements to adhere to verified reporting statements, such as Reporting

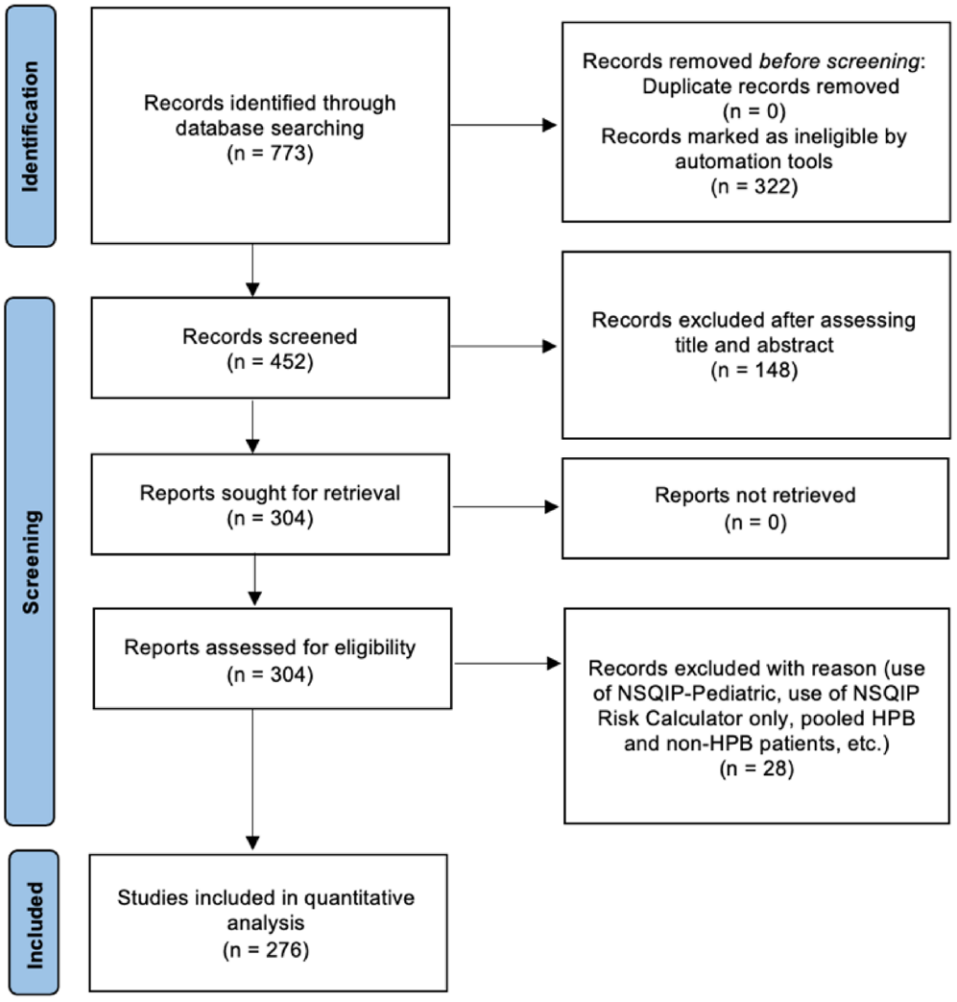

Figure 1. PRISMA diagram detailing the study selection process. HPB, hepatobiliary and pancreatic surgery.

of Studies Conducted Using Observational Routinely Collected Health Data (RECORD), STROBE, *JAMA*, PRISMA, Consolidated Standards of Reporting Trials (CONSORT), Meta-Analyses of Observational Studies in Epidemiology (MOOSE), Standards for Reporting Diagnostic Accuracy Studies (STARD), and Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT), were also documented. Critical appraisal of the final eligible publications was assessed by determining adherence to three reporting guidelines:

1) RECORD statement: The RECORD statement was first devised in 2015. Notably, the RECORD statement comprises 13 items and requires authors to provide supplementary information or accessibility to raw data. (Details of the RECORD statement are in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174.)
2) *JAMA Surgery* checklist: The editors of *JAMA Surgery* developed a 10-item checklist "to elevate the science of surgical database research" in 2018.[12] As such, this checklist aimed to guide authors in analyzing large-scale databases, such as the ACS NSQIP. (Details of the *JAMA Surgery* Checklist are in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174.)
3) STROBE statement: The STROBE statement was developed in 2007 in an effort to facilitate the reporting of observational studies by researchers. As a result of its 22 items, this tool attempts to assess the quality of writing and reporting of observational studies based on observations. Additionally, the statement emphasizes the importance of the methods used in observational studies, from the collection of data to the analysis of results. (Details of the STROBE statement are in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174.)

Each publication was scored according to the total number of fulfilled criteria with regard to the *JAMA Surgery* checklist (out of 7), STROBE statement (out of 21), and RECORD statement (out of 10). Two authors (DH and HA) independently scored each publication, and any disagreements were addressed by discussion or consultation with a third author (YK). Three criteria were excluded from the *JAMA Surgery* checklist as the evaluation was not feasible; 3 criteria from the RECORD statement were not included as they were related to database comparison; 1 criterion was excluded from the STROBE statement as it did not fall within the scope of the studies included. The criteria that were excluded from the *JAMA Surgery* checklist included those related to conducting a thorough literature review, ensuring that the data used was the best available, and ensuring that the variables used did not change over the course of the study (items 3, 4, and 7 in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174). There are 2 criteria that were excluded from the RECORD statement, which are those related to whether or not data from databases were linked to institutional level data or whether the databases were linked to each other (items 3, 6, and 10 in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174). The criterion excluded from the STROBE statement examined evolving participant eligibility (item 13 in **Supplemental Digital Content 1**, http://links.lww.com/JACS/A174). This is consistent with the methodology used by other researchers.[4] The interrater reliability was assessed by using Cohen's κ statistics across all criteria. The agreement rates for the *JAMA Surgery* Checklist, RECORD statement, and STROBE statement were 93%, 90%, and 92%, respectively. The criteria with the highest interrater reliability included: statement of compliance with IRB (*JAMA Surgery* Checklist, $\kappa = 1$), inclusion of geographic region and time frame (RECORD statement, $\kappa = 0.99$), and identification of funding sources (STROBE statement, $\kappa = 0.99$). The criteria with the lowest interrater reliability included: identification of competing risks (*JAMA Surgery* Checklist, $\kappa = 0.81$), inclusion of data cleaning methods (RECORD Statement, $\kappa = 0.81$), and specific explanation of subgroup or sensitivity analyses (STROBE Statement, $\kappa = 0.83$). Due to the nature of this study, IRB approval was not required.

Statistical analysis

Data is presented in the form of frequencies and percentages for categorical variables, while data for continuous variables is presented in the form of medians and interquartile ranges (IQRs). The first subgroup analysis compared publications in high- and low-IF journals. Journals with high IFs were considered as those scientific journals whose Journal Citation Reports indexed them in the 90th percentile relative to other surgical journals indexed in the IF.[8] A second subgroup analysis compared studies published in journals with and without a policy requiring adherence to reporting guidelines stated in the submission guidelines.[8] A third subgroup analysis compared hepatobiliary and pancreatic studies to determine if there are any differences in reporting guideline adherence based on surgical subspecialty. The Wilcoxon rank sum test was used to compare continuous variables to each other, while the Pearson chi-square test was used to compare categorical variables. Statistical significance was set at a 2-sided p Value threshold of <0.05 for all hypotheses tested. All statistical analyses were conducted using IBM SPSS statistical package (v25, IBM Corp, Armonk, NY).

RESULTS

Search results and article characteristics

Our search of the Web of Science Core Collection (all editions) generated 773 articles for possible inclusion. Among these studies, 322 were marked ineligible by automation tools, 148 were excluded after assessing the title and reading the abstract, and finally 28 were excluded as they did not meet the section criteria of the study. Included in the final analysis were 276 articles; the selection process is summarized in the PRISMA diagram shown in Figure 1. All full-text articles analyzed were published between 2006 and 2022, with the most common years of publication being 2021 (n = 47), 2019 (n = 32), and 2020 (n = 31). The most common journals of publication were *HPB* (n = 60) and the *Journal of Gastrointestinal Surgery* (n = 49). Regarding the study location, 95% of the studies (n = 265) were conducted in the US. The IF of each journal was also analyzed: 22% of the articles reviewed (n = 60) were published in high-IF journals and the median IF across all reviewed articles was 3.453 (IQR 3.267 to 3.842). Additionally, whether a journal had a reporting guideline adherence policy was also documented: 82% of articles (n = 225) were published in a journal that requires adherence to reporting guidelines (eg CONSORT, STROBE), while 18% (n = 51) were published in journals that did not include such a policy. Study characteristics and additional information are provided in Table 1.

Criteria fulfillment, according to checklist guidelines

The median scores for the *JAMA Surgery* checklist, RECORD statement, and STROBE statement were 4 of 7 (IQR 3 to 5), 3 of 10 (IQR 2 to 4), and 15 of 21

Table 1. Study Characteristics

Variable	HPB NSQIP papers (n=276)
Publication year, n (%)	
2006-2012	22 (8)
2013	12 (4)
2014	21 (8)
2015	20 (7)
2016	18 (6)
2017	28 (10)
2018	25 (9)
2019	32 (12)
2020	31 (11)
2021	47 (17)
2022	22 (8)
Journal, n (%)	
HPB	60 (22)
Journal of Gastrointestinal Surgery	49 (18)
Journal of Surgical Research	24 (9)
American Journal of Surgery	13 (5)
Journal of the American College of Surgeons	15 (5)
Surgery	19 (7)
Surgical Endoscopy and Other Interventional Techniques	17 (6)
American Surgeon	11 (4)
Annals of Surgery	11 (4)
Annals of Surgical Oncology	9 (3)
Journal of Surgical Oncology	8 (3)
Journal characteristic, n (%)	
Low IF	212 (76)
IF unknown	6 (2)
Submission guidelines require adherence to reporting guidelines (eg CONSORT, STROBE, etc)	225 (82)
Submission guidelines do not require adherence to reporting guidelines (eg CONSORT, STROBE, etc)	51 (18)

CONSORT, Consolidated Standards of Reporting Trials; IF, impact factor; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

(IQR 13 to 17), respectively. Regarding the *JAMA Surgery* checklist, the majority of studies demonstrated a clear take-home message (n = 268) and identified potential confounders (n = 248). Additionally, 175 studies included a statement of compliance with an institutional review board, 157 studies addressed data issues, and 92 studies proposed a solid research question and hypothesis. Few studies addressed competing risks (n = 19), and only 50 studies utilized a flow diagram to describe the final patient selection. In examining the RECORD statement checklist, all studies specified the type of data used (n = 276), which was expected given that all publications examined were required to utilize the ACS NSQIP database. The majority of studies clearly stated the population selection and included specific CPT or ICD codes used for data acquisition (n = 195). Additionally, 159 studies included information about geographic region and timeframe; 116 studies discussed exposures, outcomes, confounders, and effect modification; and 82 studies discussed the implications of using a dataset that was not created to answer the authors' specific research question. A minority of studies included information on database access (n = 77) and 61 studies described how the algorithms used for participant selection were validated, with only 26 studies addressed data cleaning methods. Very few studies included instructions on accessing supplemental information (n = 25) and even fewer specifically discussed data quality and availability in the context of participant selection (n = 23). In the STROBE checklist, all studies included a well-organized abstract that summarized the study findings, provided background on the purpose of the study, detailed the study design, reported outcome data of interest, and summarized key results (n = 276). Only a minority of the studies addressed the generalizability of the results (n = 65), and many studies did not include subgroups or sensitivity analyses (n = 72). The remaining fulfillment criteria are detailed in Table 2.

High-IF vs low-IF checklist fulfillment

Of the 276 included studies, 60 were published in a high-IF journal, 210 were published in a low-IF journal, and 6 were published in journals with an unknown IF and therefore were not included in this analysis. In comparing these subgroups, there was no statistically significant difference in criteria fulfillment of the *JAMA Surgery* checklist. However, in terms of the RECORD statement checklist fulfillment, studies published in high-IF journals were significantly more likely to discuss exposures, outcomes, confounders, and effect modification compared to low-IF journals (55% vs 38%, $p = 0.019$). In examining the STROBE statement checklist fulfillment, studies published in high-IF journals were significantly more likely to identify specific objectives and include a predetermined hypothesis compared with studies published in low-IF journals (33% vs 7%, $p < 0.0001$). Studies published in high-IF journals addressed the generalizability of the study more often than those published in low-IF journals (33% vs 21%, $p = 0.047$). Since the RECORD statement and the STROBE statement have overlapping checklist items (discussion of exposures, outcomes, confounders, and effect modification), it is expected that this finding would be significant in examining both checklists. Additional

Table 2. Checklist Item Fulfillment

Checklist item	Data, n (%)
JAMA Surgery checklist*	
The article has a solid research question and clear hypothesis	92 (33)
Ensure compliance with the institutional review board and data use agreements	175 (63)
Clearly define the inclusion criteria, exclusion criteria, and outcome variables using a flow diagram to describe final patient selection	50 (18)
Identify potential confounders and use risk adjustment to minimize bias, avoiding use of causal language in reporting results of these observational studies.	248 (89)
Ensure that competing risks are identified and addressed	19 (7)
Ensure that data issues, such as missing data, are discussed and that any sensitivity analyses or imputations performed are reported in a clear and cohesive way	157 (56)
The article has a clear take-home message that addresses how this research advances current knowledge and has important policy or clinical implications	268 (96)
RECORD statement	
Type of data specified	276 (100)
Geographic region and timeframe identified	159 (57)
Study population selection methodology (codes, algorithms)	195 (70)
Validation studies	61 (22)
Exposures, outcomes, confounders, and effect modifiers included	116 (42)
Database access	77 (28)
Data cleaning methods addressed	27 (10)
Study population selection (data quality, availability)	23 (8)
Implications of using data not created for the specific research question(s)	82 (29)
Supplemental information access	25 (9)
STROBE statement	
Title and abstract: balanced summary of study design	276 (100)
Introduction	
Background and scientific rationale	276 (100)
Specific objectives and prespecified hypotheses	92 (33)
Methods	
Key elements of study design	276 (100)
Setting, location, and relevant dates	159 (58)
Eligibility criteria, sources, and methods of participant selection	114 (41)
Outcomes, exposures, potential confounders, and effect modifiers defined	116 (42)
Data sources and methods of assessment	195 (71)
Potential sources of bias addressed	143 (52)
Study size explained	189 (68)
Quantitative variable analysis	256 (93)
Statistical methods, subgroup analysis, and missing data described	157 (57)
Results	
Participant characteristics and information on exposures	76 (28)
Reporting of outcome data	276 (100)
Confounder-adjusted estimates and their precision	212 (77)
Subgroup and sensitivity analyses	72 (26)
Discussion	
Summary of key results	276 (100)
Limitations of the study are discussed	242 (88)

(Continued)

Table 2. Continued

Checklist item	Data, n (%)
Overall interpretation of results, considering limitations and results from similar studies	250 (91)
External validity of study results	65 (11)
Other: funding	95 (34)

RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.
*Reprinted from Haider and colleagues,[12] with permission from *JAMA Surgery*.

information and statistical significance levels are presented in Table 3.

Checklist fulfillment before vs after checklist publication

The RECORD statement was first published in October 2015, whereas the *JAMA Surgery* checklist was first published in June 2018, so we sought to analyze whether checklist item fulfillment increased as a result of checklist publication. Since the STROBE statement was first published in 2007, these data were not included in this analysis, as all studies analyzed were published after the introduction of the STROBE statement, except for 1 study published in 2006. After the publication of the *JAMA Surgery* checklist, there was a significant increase in IRB compliance statements (74% vs 52%, p = 0.0001) and inclusion of a clear definition of inclusion/exclusion criteria (24% vs 11%, p = 0.005). Interestingly, this analysis also revealed a significant decrease in studies that addressed competing risks (3% vs 11%, p = 0.037). (See Table 4 for additional information.) After publication of the RECORD statement, there was a significant increase in the number of studies that specifically described exposures, outcomes, confounders, and effect modification (48% vs 25%, p = 0.001). Unexpectedly, studies were less likely to include information regarding geographic region and study timeframe (52% vs 74%, p = 0.002) after the introduction of the RECORD statement. (See Table 5 for additional information). Though the publication of the *JAMA Surgery* and RECORD statement checklists impacted individual checklist item fulfillment, there was no significant change in the mean number of checklist items fulfilled per year, as shown in Figure 2.

Effect of journal requirement of adherence to reporting standards

We also performed a subgroup analysis comparing articles published in journals with and without a reporting guideline adherence policy. The publications examined in the current study were published in a variety of journals: 82% of studies (n = 225) were published in journals that mandate adherence to reporting guidelines, while 18% of studies (n = 52) were published in journals that do not state this requirement. We then compared the mean checklist item fulfillment between these 2 subgroups, as shown in Figure 3. While there was no significant difference found between the 2 groups, there is a trend towards increased checklist item fulfillment noted in favor of journals that have a reporting guideline adherence policy.

Surgical subspecialty: hepatobiliary vs pancreatic surgery publications

We performed a third subgroup analysis comparing the checklist item fulfillment of hepatobiliary and pancreatic surgery publications. Ten articles were not included in this analysis as they examined hepatobiliary and pancreatic surgery outcomes collectively. While there was no significant difference in the median scores of the *JAMA Surgery* and RECORD statement checklists, pancreatic surgery publications had significantly improved fulfillment of the STROBE statement checklist compared to hepatobiliary surgery publications (p = 0.0002). Median scores and interquartile ranges are presented in Table 6.

DISCUSSION

The results indicate that there were many HPB publications that met the criteria outlined by the 3 established reporting guidelines (*JAMA Surgery*, STROBE, and RECORD). As a result, data quality and availability need to be discussed, competing risks should be acknowledged, missing data should be evaluated, supplementary/raw information should be provided, adjusted and unadjusted estimates/outcomes should be reported, and subgroup analysis should be performed.[9] This study, therefore, calls for enhancing the quality of HPB studies that utilize large databases. Interestingly, analysis of how checklist publication impacted individual item fulfillment was inconclusive, as some items were fulfilled significantly more often after publication, while others were fulfilled significantly less often. This finding prompts further investigation into the impact of the publication of these checklists on the methodological quality of research more broadly. Additionally, in examining the effect of

Table 3. Checklist Item Fulfilment, High-Impact Journal Publications vs Low-Impact Journal Publications

Checklist item fulfillment, n (%)	High IF Journal (n = 60)	Low IF Journal (n = 210)	p Value
JAMA Surgery checklist*			
Solid research question and clear hypothesis	17 (28)	75 (35)	0.309
Statement of IRB compliance	33 (55)	136 (64)	0.174
Clear definition of inclusion/exclusion criteria, outcome variables, and inclusion of a flowchart diagram	8 (13)	41 (19)	0.285
Identified potential confounders and discussed procedures to minimize bias	56 (93)	186 (88)	0.256
Identified and addressed competing risks	2 (3)	17 (8)	0.209
Discussed data issues, such as missing data	36 (60)	119 (56)	0.593
Clear take-home message regarding advancement of clinical knowledge and implications	58 (97)	205 (97)	1
RECORD statement			
Type of data specified	60 (100)	210 (100)	0.653
Geographic region and timeframe identified	32 (53)	124 (58)	0.476
Study population selection methodology (codes, algorithms)	44 (73)	145 (68)	0.463
Validation studies	9 (15)	51 (24)	0.135
Exposures, outcomes, confounders, and effect modifiers included	33 (55)	80 (38)	0.019†
Database access	19 (32)	57 (27)	0.466
Data cleaning methods addressed	7 (12)	20 (9)	0.610
Study population selection (data quality, availability)	4 (7)	19 (9)	0.573
Implications of using data not created for the specific research question(s)	12 (20)	69 (33)	0.061
Supplemental information access	5 (8)	20 (9)	0.794
STROBE statement			
Title and abstract: balanced summary of study design	60 (100)	210 (100)	1
Introduction			
Background and scientific rationale	60 (100)	210 (100)	1
Specific objectives and prespecified hypotheses	20 (33)	14 (7)	<0.0001†
Methods			
Key elements of study design	60 (100)	210 (100)	1
Setting, location, and relevant dates	32 (53)	124 (59)	0.476
Eligibility criteria, sources, and methods of participant selection	22 (37)	87 (41)	0.507
Outcomes, exposures, potential confounders, and effect modifiers defined	33 (55)	80 (38)	0.019†
Data sources and methods of assessment	47 (78)	145 (69)	0.162
Potential sources of bias addressed	25 (42)	114 (54)	0.085
Study size explained	43 (72)	141 (67)	0.507
Quantitative variable analysis	57 (95)	203 (97)	0.364
Statistical methods, subgroup analysis, and missing data described	39 (65)	114 (54)	0.140
Results			
Participant characteristics and information on exposures	20 (33)	55 (26)	0.276
Reporting of outcome data	60 (100)	210 (100)	1
Confounder-adjusted estimates and their precision	44 (73)	162 (77)	0.541
Subgroup and sensitivity analyses	20 (33)	51 (24)	0.160
Discussion			
Summary of key results	60 (100)	210 (100)	1
Limitations of the study are discussed	52 (87)	184 (88)	0.845
Overall interpretation of results, considering limitations and results from similar studies	56 (93)	188 (90)	0.378

(Continued)

Table 3. Continued

Checklist item fulfillment, n (%)	High IF Journal (n = 60)	Low IF Journal (n = 210)	p Value
External validity of study results	20 (33)	44 (21)	0.047†
Other: funding	21 (35)	72 (34)	0.918

†p Value <0.05.
IF, impact factor; RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.
*Reprinted from Haider and colleagues[12] with permission from *JAMA Surgery*.

Table 4. *JAMA Surgery* Checklist Fulfilment: Before and after Checklist Publication

Checklist item, n (%)	2006 to June 2018 (n = 132)	June 2018 to 2022 (n = 144)	p Value
Solid research question and clear hypothesis	47 (36)	45 (31)	0.443
Statement of IRB compliance	69 (52)	107 (74)	0.0001*
Clear definition of inclusion/exclusion criteria, outcome variables, and inclusion of a flowchart diagram	15 (11)	35 (24)	0.005*
Identified potential confounders and discussed procedures to minimize bias	115 (87)	134 (93)	0.097
Identified and addressed competing risks	14 (11)	5 (3)	0.037*
Discussed data issues, such as missing data	71 (54)	87 (60)	0.266
Clear take-home message regarding advancement of clinical knowledge and implications	130 (98)	139 (97)	0.302

*p Value <0.05.
Reprinted from Haider and colleagues[12] with permission from *JAMA Surgery*.

Table 5. RECORD Statement Checklist Fulfilment: Before and after Checklist Publication

Checklist item, n (%)	2006 to Oct 2015 (n = 68)	Oct 2015 to 2022 (n = 208)	p Value
Type of data specified	68 (100)	208 (100)	1
Geographic region and timeframe identified	50 (74)	109 (52)	0.002*
Study population selection methodology (codes, algorithms)	52 (76)	143 (69)	0.225
Validation studies	10 (15)	51 (25)	0.090
Exposures, outcomes, confounders, and effect modifiers included	17 (25)	99 (48)	0.001*
Database access	18 (26)	59 (28)	0.762
Data cleaning methods addressed	7 (10)	20 (10)	0.870
Study population selection (data quality, availability)	5 (7)	18 (9)	0.736
Implications of using data not created for the specific research question(s)	15 (22)	67 (32)	0.112
Supplemental information access	5 (7)	20 (10)	0.573

*p value <0.05.
RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data.

reporting guideline adherence policies, there was no statistical difference in mean checklist item fulfillment when comparing journals with and without such policies. As such, this raises questions regarding the optimization of reporting guideline requirements and the possible need for stricter adherence policies. With an increasing number of studies using the ACS NSQIP database published every year, it is imperative that the standards of quality research be maintained to ensure accurate and accessible reporting.

Several studies have been performed on the methodological quality of ACS NSQIP studies. For example, in 2020, Yolcu and associates examined neurosurgical ACS NSQIP publications and scored them according to the 3 established reporting guidelines used in this study. Similar to our research, the most commonly unfulfilled criterion was the inclusion of supplementary information.[7] El Moheb and associates assessed the methodologic reporting of emergency general surgery studies,[8] and found numerous studies that did not explain the

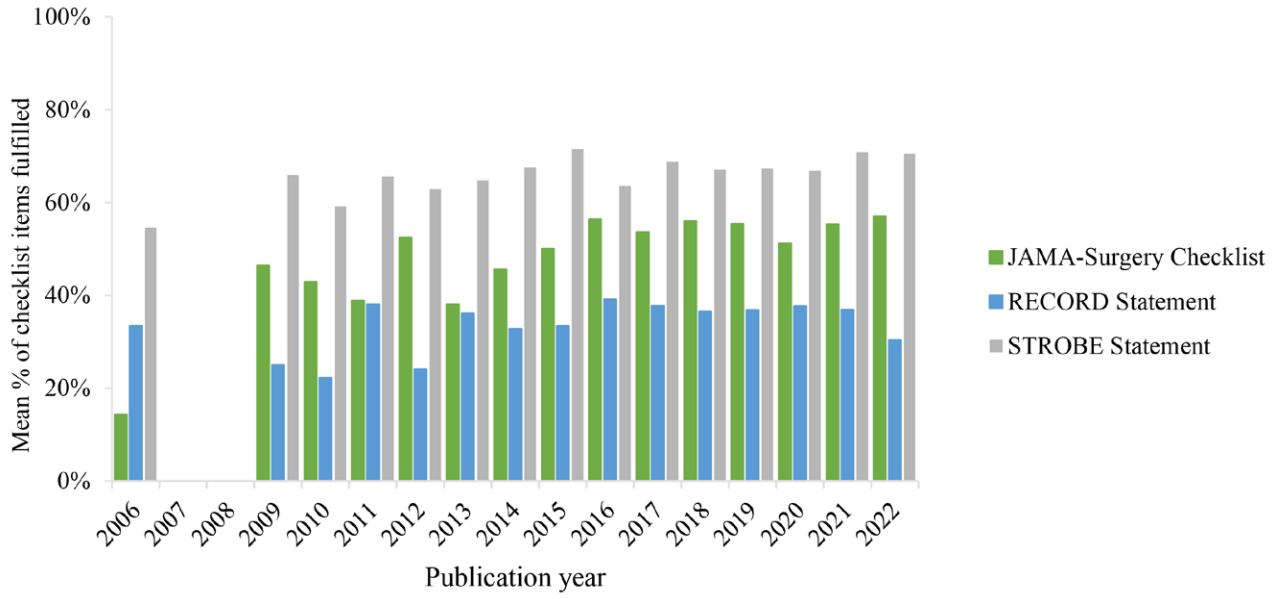

Figure 2. Mean percentage of checklist items fulfilled over time. There is no significant change in the percentage of checklist items fulfilled before and after publication of reporting guidelines. RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

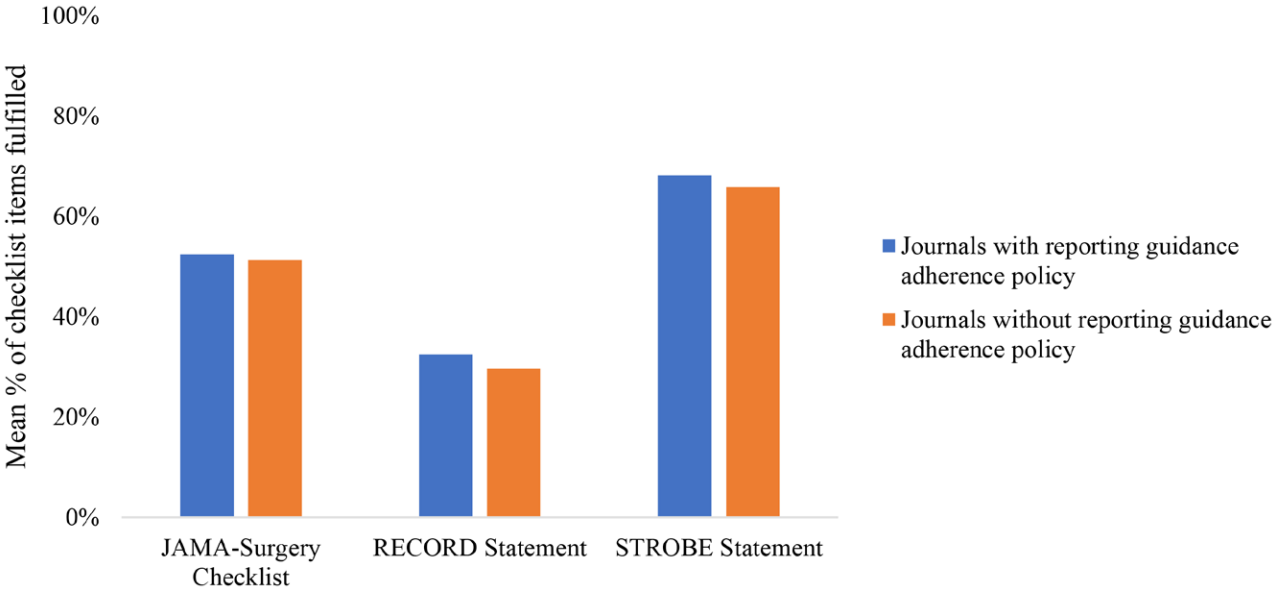

Figure 3. Mean percentage of checklist item fulfillment is similar in journals with reporting guideline adherence requirements vs journals without reporting guideline adherence requirements. RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

methodology for data cleaning, population selection process, or management of missing variables. This has resulted in an increasing body of literature that critically scrutinizes studies utilizing large-scale databases in the light of these initiatives. Recently, Oravec and associates published a paper that recommended that researchers should consider certain limitations before proceeding with a study that involves large-scale data in order to ensure the results are relevant and accurate.[13] Our study showed that most HPB studies reported limitations, but

Table 6. Checklist Item Fulfilment: Pancreatic vs Hepatobiliary Studies

Variable	Pancreatic studies*	Hepatobiliary studies*	p Value
JAMA Surgery checklist	4 (5-3)	4 (4-3)	0.082
RECORD statement	3 (4-2)	3 (4-2)	0.395
STROBE statement	15 (17-14)	14 (16-13)	0.0002†

*Data presented as median (interquartile range).
†p Value <0.05.
Ten studies were not included in this subgroup analysis as they investigated hepatopancreatic-biliary outcomes collectively.
RECORD, Reporting of Studies Conducted Using Observational Routinely Collected Health Data; STROBE, Strengthening the Reporting of Observational Studies in Epidemiology.

many failed to address bias, competing risks, and external validity. Khachfe and associates evaluated ACS NSQIP studies using the pancreas database and found that the median scores for several criteria for the *JAMA Surgery* checklist, STROBE statement, and RECORD statement were 6, 20, and 6, respectively.[4] There were several criterion not met in the majority of cases, including discussing missed data, compliance with the IRB, providing unadjusted and adjusted outcomes, and performing subgroup analyses. The current study advances the findings presented by Khachfe and associates by examining hepatobiliary surgery publications and pancreatic surgery publications separately; our analysis revealed that pancreatic surgery publications demonstrated greater adherence to the STROBE statement guidelines. As such, this finding should serve as a call to action among researchers to specifically improve the quality of methodological reporting in the field of hepatobiliary surgery. In summary, there is a lack of consistency across specialties in the statistical analysis of surgical outcome data and in addressing the potential pitfalls of using large-scale databases, such as ACS NSQIP.

One limitation of utilizing nationwide datasets is the balance between clinical and statistical significance. Although this limitation may seem concerning, it is reassuring that authors of HPB publications using these databases do seem to understand the need for reporting limitations, regardless of the journal of publication. Studies in high-IF journals had more reporting related to documenting specific objectives and predetermined hypotheses, assessing external validity, and identifying exposures, outcomes, confounders, and effect modifiers.[14] However, most checklist items did not reveal a statistically significant difference in fulfillment when comparing high- vs low-IF subgroups. Looking to other specialties, Yolcu and associates similarly found that many neurosurgery publications did not mention potential sources of bias, data cleaning methods, supplemental information, or external validity. They also found fulfillment rates were not significantly impacted by journal IF.[7]

Despite the fact that HPB publications using the ACS NSQIP database are generally compliant with methodologic reporting guidelines, there is still room for improvements to be made in this aspect. By adhering to guidelines, surgical research can be optimized and higher-quality publications can be produced.[15-17] We suggest that a checklist be incorporated into the manuscript submission process to ensure that all guidelines are met before editorial review. Additionally, it is essential that concerted efforts by organizations, such as the International Committee of Medical Journal Editors, should be implemented in order to enforce stricter adherence to these standards. Moreover, in spite of the fact that the *JAMA Surgery* checklist, STROBE statement, and RECORD statement are widely recognized appraisal tools, they are all limited by their lack of specificity.[16-20] A specialty-specific checklist, drafted by HPB surgeons who are experienced in the field, could be helpful in elevating the quality of HPB research. If this were to happen, surgical research would be more consistent and less likely to be underreported. In future studies, it will be important to investigate the impact of new appraisal instruments on the quality of the studies over an extended period of time as well.

These findings and analyses should be interpreted in the context of this study's limitations. First, we scored items on the *JAMA Surgery* checklist, the STROBE statement, and the RECORD statement as binary. Although this method is necessary to provide objective quantitative assessments, the checklist items are not necessarily of equal importance. Second, increased adherence to standardized guidelines does not necessarily translate to enhanced research quality, though it does serve as a useful objective marker of study design. It is also important to note that certain criteria were excluded in the current study because they were not relevant to the specific subset of publications examined. This practice may raise concern that researchers similarly overlook criteria they have deemed irrelevant or out of the scope of their investigation. This customizability generates a potentially problematic extension of standardized reporting guidelines. Despite these limitations, we generally believe that adherence to the reporting guidelines can help improve the quality of surgical

studies and provide an essential foundation for the production of valuable research. Of final note, rejected manuscripts and in-press publications were not available for analysis, so we could only assess published articles. These manuscripts may have held different findings, such as adequate or suboptimal reporting.

CONCLUSION

ACS NSQIP provides a unique opportunity for researchers to access large-scale, nationwide data; however, caution must be exercised to maintain adherence to surgical research reporting guidelines. While HPB studies using ACS NSQIP fulfill many of the criteria put forth by the *JAMA Surgery* checklist, RECORD statement, and STROBE statement, there are areas for improvement, including discussing competing risks, providing supplementary information, and performing subgroup analysis. Finally, this study calls for further investigation into the efficacy of reporting guidelines and potential development of new tools that would enhance the methodologic quality of HPB research.

Author Contributions

Data curation: Hockett, Rabinowitz
Formal analysis: Hockett, Rabinowitz
Methodology: Hockett, Rabinowitz, Kwon
Investigation: Rabinowitz
Writing – review & editing: Rabinowitz, Kaafarani, Aziz
Project administration: Kwon
Resources: Kwon
Software: Joseph
Visualization: Joseph
Writing – original draft: Joseph, Kaafarani, Aziz
Conceptualization: Kaafarani, Aziz
Validation: Kaafarani

REFERENCES

1. Beyranvand T, Aryankhesal A, Aghaei Hashjin A. Quality improvement in hospitals' surgery-related processes: A systematic review. Med J Islam Repub Iran 2019;33:129.
2. Molina CS, Thakore RV, Blumer A, et al. Use of the National Surgical Quality Improvement Program in orthopaedic surgery. Clin Orthop Relat Res 2015;473:1574e–11581.
3. ACS American College of Surgeons. National Surgical Quality Improvement Program. Available at: https://www.facs.org/Quality-Programs/ACS-NSQIP. Accessed June 2022.
4. Khachfe HH, Habib JR, Salhab HA, et al. American College of Surgeons NSQIP pancreatic surgery publications: A critical appraisal of the quality of methodological reporting. Am J Surg 2022;223:705–714.
5. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. J Am Coll Surg 2013;217:833833e–83842e3.
6. Aziz H, Hanna K, Saif MW, et al. Hepatectomy for breast cancer metastasis and sarcoma are more likely to have adverse outcomes than hepatectomy for primary hepatocellular cancer or for colorectal metastasis. Cancer Med J 2020;3:6–12.
7. Yolcu Y, Wahood W, Alvi MA, et al. Reporting methodology of neurosurgical studies utilizing the American College of Surgeons-National Surgical Quality Improvement Program Database: A systematic review and critical appraisal. Neurosurgery 2020;86:46–60.
8. El Moheb M, Sabbagh H, Badin D, et al. Appraising the quality of reporting of American College of Surgeons NSQIP emergency general surgery studies. J Am Coll Surg 2021;232:671–680.
9. Gebran A, Bejjani A, Badin D, et al. Critically appraising the quality of reporting of American College of Surgeons TQIP studies in the era of large data research. J Am Coll Surg 2022;234:989–998.
10. Aziz H, Kwon YIC, Alvi S, et al. Does chronic use of steroids affect outcomes after liver resection? Analysis of a National Database [published online ahead of print, July 1, 2022]. J Gastrointest Surg 2022;26:2093–2100.
11. Clarivate. Journal citation reports. Available at: https://jcr.clarivate.com/. Accessed July 2022.
12. Haider AH, Bilimoria KY, Kibbe MR. A checklist to elevate the science of surgical database research. JAMA Surg 2018;153:505e–5507.
13. Oravec CS, Motiwala M, Reed K, et al. Big data research in neurosurgery: A critical look at this popular new study design. Neurosurgery 2018;82:728–746.
14. Blanco D, Kirkham JJ, Altman DG, et al. Interventions to improve adherence to reporting guidelines in health research: A scoping review protocol. BMJ Open 2017;7:e017551e017551.
15. Englesbe MJ, Pelletier SJ, Magee JC, et al. Seasonal variation in surgical outcomes as measured by the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP). Ann Surg 2007;246:456e–4462.
16. Khera R, Angraal S, Couch T, et al. Adherence to methodological standards in research using the National Inpatient Sample. JAMA 2017;318:2011–2018.
17. Bridoux V, Moutel G, Roman H, et al. Methodological and ethical quality of randomized controlled clinical trials in gastrointestinal surgery. J Gastrointest Surg 2012;16:1758–1767.
18. Ahmed Ali U, Reiber BMM, Ten Hove JR, et al. Journal impact factor and methodological quality of surgical randomized controlled trials: An empirical study. Langenbecks Arch Surg 2017;402:1015–1022.
19. Balasubramanian SP, Wiener M, Alshameeri Z, et al. Standards of reporting of randomized controlled trials in general surgery: Can we do better? Ann Surg 2006;244:663–667.
20. Pussegoda K, Turner L, Garritty C, et al. Systematic review adherence to methodological or reporting quality. Syst Rev 2017;6:131.