Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Assessing quality of randomized trials supporting guidelines for laparoscopic and endoscopic surgery

Aaron Bowers, BS,[a,*] *Chase Meyer, BS,*[a] *Daniel Tritz, BS,*[a] *Courtney Cook, BS,*[a] *Kaleb Fuller, BS,*[a] *Caleb Smith,*[a] *Brian Diener, DO,*[b] *and Matt Vassar, PhD*[a]

[a] *College of Osteopathic Medicine, Oklahoma State University Center for Health Sciences, Tulsa, Oklahoma*
[b] *Department of Surgery, Oklahoma State University Medical Center, Tulsa, Oklahoma*

ARTICLE INFO

Article history:
Received 5 August 2017
Received in revised form
27 October 2017
Accepted 22 November 2017
Available online 22 February 2018

Keywords:
Fragility
Risk of bias
Guidelines
Surgery
Randomized controlled trial

ABSTRACT

Background: Recent studies have highlighted the risk of bias and the fragility of results in randomized controlled trials (RCTs). The aim of our study was to evaluate the clinical practice guidelines created by the Society for Gastrointestinal and Endoscopic Surgeons (SAGES) for fragility, statistical power, and risk of bias.

Materials and methods: We screened the SAGES clinical practice guideline references for qualifying RCTs. RCTs were assessed for risk of bias using the Cochrane Collaboration Risk of Bias tool 2.0. We used the fragility index and fragility quotient to evaluate the robustness of trial results and conducted a power analysis using G*Power to determine if trials were adequately powered.

Results: Twenty-two (40.7%) of the 54 trials that we assessed were rated as having a high risk of bias, 17 (31.5%) were rated as having a low risk of bias, and 15 (27.8%) were rated as having some concerns. The median fragility index was 2.5 (interquartile range 1-7). The median fragility quotient was 0.021 (interquartile range 0.003-0.045). Mean sample size was 108, and the mean loss to follow-up was eight patients. Eight of 33 trials (24.2%) were found to be underpowered according to the sample size used in the primary outcome.

Conclusions: Guidelines created by SAGES are supported by RCTs that are frequently fragile or underpowered or have a high risk of bias. Future RCTs should utilize the Consolidated Standards of Reporting Trials statement, implement strategies to minimize loss to follow-up, and use properly powered sample sizes.

Introduction

An estimated 3.5 million laparoscopic surgeries are performed annually worldwide.[1] These surgeries are used to treat a variety of conditions including colon cancer, hernias, gastroesophageal reflux disease, biliary tract diseases, appendicitis, and obesity. From 2009 to 2014, the number of laparoscopic appendectomies, colectomies, and hernia repairs more than doubled.[2] Given the increasing frequency of laparoscopic procedures, it is important for surgeons to

All authors declare they have no conflicts of interest.

* *Corresponding author.* Mr Aaron Bowers, Oklahoma State University Center for Health Sciences, 1111 W 17th Street, Tulsa, OK 74107. Tel.: +1 801 808 0187; fax: +1 918 582 1972.
E-mail address: aaron.bowers@okstate.edu (A. Bowers).

0022-4804/$ – see front matter © 2017 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.jss.2017.11.061

incorporate the best available evidence when caring for patients.

The Society for Gastrointestinal and Endoscopic Surgeons (SAGES) has produced 17 clinical practice guidelines (CPGs) to inform surgeons regarding the management of some of the most common diseases in medicine, with treatments including biliary tract surgery, appendectomy, treatment of colon cancer, and the treatment of gastroesophageal reflux disorder. These guidelines are based on a comprehensive review of evidence and developed by a panel of experts. Recommendations within the guidelines are used to direct evidence-based patient care. Several researchers, however, have called into question the effectiveness of CPGs for improvement in patient outcomes, finding that patients managed according to CPG recommendations have only minimal improvement compared with patients managed without CPG recommendations.[3,4] Many CPGs, including those produced by SAGES, use the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) system for rating the level of evidence supporting their recommendations. The GRADE system regards statistically significant differences from randomized controlled trials (RCTs) as the highest quality of evidence.[5]

Results from RCTs can be more confidently attributed to an intervention than results based on other study designs because potential sources of bias are minimized through methodological safeguards such as randomization and blinding.[6] Recent studies, however, have called into question the robustness, or strength, of RCT results and the likelihood of bias in these trials due to flawed study designs.[7-10] Bias in RCT execution or reporting has the potential to result in inappropriate management decisions. Ioannidis reported that among interventions with superior results from clinical trials, 32% were later found to be harmful or to have no effect.[11] The quality of RCTs is critical because the evidence they provide constitutes the basis for selecting the most appropriate therapy.

To improve CPGs and patient outcomes, the RCTs supporting them need to be well-conducted and well-reported and to have an adequate sample size. The aim of our study was to objectively evaluate the RCTs supporting the SAGES guidelines because no prior assessment of the RCTs has been carried out. In our assessment, we determined the fragility of RCTs underpinning some guideline recommendations, evaluated whether these RCTs were adequately powered, and assessed the risk of bias of these trials.

Methods

Eligibility criteria

We performed a systematic survey of all RCTs referenced in the following 11 SAGES CPGs: Guidelines for Laparoscopic Ventral Hernia Repair; Guidelines for Laparoscopic Peritoneal Dialysis Access Surgery; Guidelines for the Management of Hiatal Hernia; Guidelines for Laparoscopic Resection of Curable Colon and Rectal Cancer; Guidelines for Diagnosis, Treatment, and Use of Laparoscopy for Surgical Problems During Pregnancy; Guidelines for Surgical Treatment of Gastroesophageal Reflux Disease; Guidelines for the Clinical Application of Laparoscopic Biliary Tract Surgery; Guidelines for Laparoscopic Appendectomy; Guidelines for the Surgical Treatment of Esophageal Achalasia; Guidelines for Diagnostic Laparoscopy; and Guidelines for Clinical Application of Laparoscopic Bariatric Surgery.[12] Five CPGs were excluded from analysis: Guidelines for Office Endoscopic Services and Guidelines for the Surgical Practice of Telemedicine because they lack recommendations; Guidelines for the Minimally Invasive Treatment of Adrenal Pathology and Guidelines for the Use of Laparoscopic Ultrasound because they lack usable RCTs; and venous thromboembolism Prophylaxis for Laparoscopy Surgery Guidelines: An update because it serves as an endorsement of a guideline put forward by the American College of Chest Physicians. Identified RCTs were then evaluated according to an inclusion criteria determined *a priori*. For fragility calculations, a trial must have reported the random assignment of patients to a condition using a 1:1 allocation ratio, a parallel two-group design, and one or more dichotomous outcomes.

Identification of studies and data collection

We surveyed the SAGES CPGs to identify all possible RCTs both referenced and cited within the document. We searched PubMed on July 10, 2017, to obtain the abstracts and full-text articles of the eligible RCTs. Two investigators (A.B. and C.M.) surveyed the SAGES guidelines for eligible studies, reviewed the abstracts, and performed a complete screening of included trials. Duplicates were removed.

Data were derived from each of the included randomized trial using piloted electronic forms. The data included the total sample size, the sample size of each group, the number lost to follow-up for each group (if reported), the outcome reported, event rates for the outcome, statistical significance for the included outcome, and the statistical test used for comparison. When available, we used the outcome that was given in the guideline recommendations. In the cases in which this approach was not possible, the primary outcome was prioritized for analysis; otherwise, we used secondary or unspecified dichotomous outcomes. If a trial had multiple dichotomous outcomes, we used the GRADE Network's[13] approach to select the most important outcome. For this approach, a board-certified general surgeon (B.D.) was consulted. Outcomes were ranked according to importance, and the highest ranked outcome was included for analysis.

Fragility index and the fragility quotient

Fragility is a measurement of the robustness of statistically significant results. The fragility index (FI) is a number equal to the number of individuals who would have to switch from an event to a nonevent in order for the results to lose statistical significance. For example, suppose a trial has two arms, each with 50 individuals. In the first group, 18 people experience a given outcome. Eight experience the same outcome in the second group, yielding a P value of 0.039. However, if one more person in the second group achieved the outcome, or if one fewer person in the first group achieved the outcome, the P value goes down to 0.07. Therefore, the FI in this trial is one.