Original Research

An Estimate of Severe Harms Due to Screening Colonoscopy: A Systematic Review

Alison N. Huffstetler, MD, Joseph Fraiman, MD, Shannon Brownlee, MSc, Michael A. Stoto, PhD, and Kenneth W. Lin, MD, MPH

Objective: This study aims to comprehensively assess the direct, severe harms of screening colonoscopy in the United States. Whereas other investigators have completed systematic reviews estimating the harms of all types of colonoscopy, this analysis focuses on screening colonoscopies that had adequate follow up to avoid undercounting delayed harms.

Data Sources: PubMed and Embase were queried for relevant studies on screening colonoscopy harms published between January 1, 2002, and April 1, 2022.

Study Selection: English-language studies of screening colonoscopy for average risk patients were included. Studies must have followed patients for adequate time post procedure, defined as 30 days after colonoscopy.

Main Outcomes: The primary outcome was the number of severe bleeding events and gastrointestinal (GI) perforations within 30 days of screening colonoscopy.

Results: A total of 1951 studies were reviewed for inclusion; 94 were reviewed in full text. Of those reviewed in full, 6 studies, including a total of 467,139 colonoscopies, met our inclusion criteria and were included in our analysis of harms related to screening colonoscopies. The rate of severe bleeding ranged credibly from 16.4 to 36.18 per 10,000 colonoscopies; the rate of perforation ranged credibly from 7.62 to 8.50 per 10,000 colonoscopies.

Conclusions: This study is the first to estimate direct harms from screening colonoscopy, including harms that occur up to 30 days after the procedure. The risk of harm subsequent to screening colonoscopy is higher than previously reported and should be discussed with patients when engaging in shared decision making. (J Am Board Fam Med 2023;00:000–000.)

Keywords: Colonoscopy, Colorectal Cancer, Early Detection of Cancer, Long Term Adverse Effects, Patient Harm, Preventive Medicine, Systematic Review

Introduction

For more than 2 decades, the United States Preventive Services Task Force (USPSTF) and other national organizations have recommended screening for colorectal cancer, and colonoscopy being 1 of 4 options for colorectal cancer screening purposes. Although serious harm is known to occur secondary to screening colonoscopy, the USPSTF concluded the risk of harm is outweighed by potential benefits, depending on the patient’s age, risk factors and prior colonoscopic findings, provided the colonoscopy is performed at the recommended schedule for average-risk adults between 45 and 75 years old. Guidelines issued by the US Multi-Society Task Force are similar to those of the USPSTF. Complications associated with screening colonoscopy can be due to bowel preparation, anesthesia, and...
and the procedure itself.\textsuperscript{4,5} Severe harms include bowel perforation and bleeding requiring hospitalization, as well as infection, cardiovascular events, and death. Such severe harms are reducible with good colonoscopic technique, but not completely avoidable.

In recent years, the recognition that some proportion of patient harm is inescapable has increased attention on the concept of “preventable harm.”\textsuperscript{6} At the same time there has been a growing recognition that “low-value care,” or medical services that are unnecessary or inappropriate, should be viewed as a source of preventable patient harm.\textsuperscript{6} The authors conducted this systematic review of the annual rate of serious harms secondary to screening colonoscopy with the intention of using the results for a future study, which will estimate rates of preventable harm from inappropriate colonoscopy. This study differs from prior systematic reviews in 2 ways: it aims to evaluate harms exclusively in the setting of screening, as opposed to diagnostic colonoscopy, and it includes harms that commonly develop days to weeks after the procedure.

\section*{Methods}
We conducted a systematic review of the serious harms of colonoscopy. Eligible trials for harms were identified by searches of PubMed and Embase. Studies published between January 1, 2002 and April 1, 2022 were included in the searches, to include only those colonoscopies that took place after the 2002 initial primary prevention USPSTF recommendation for primary screening via colonoscopy. Search strategies for harms are included in Supplement 1. We followed the statement on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for RCTs\textsuperscript{7} (Figure 1) and this study is listed under PROSPERO record CRD42019117883.

\subsection*{Eligibility Criteria for Systematic Reviews}
Studies evaluating the harms of colonoscopy were considered eligible if the study: (1) was a prospective or retrospective analysis of the harm of colonoscopy; (2) included perforation and severe bleeding; (3) included 30 days of follow up; (4) excluded patients for whom colonoscopy was likely for diagnostic purposes rather than screening, for example, surveillance colonoscopies, high risk patients, patients with symptoms of colorectal cancer, or those with a personal history of inflammatory bowel disease or colorectal cancer; (5) did not duplicate the data source of another, more robust study.

\subsection*{Study Identification}
Two independent investigators (J.F. and A.H.) screened articles by title and abstract using standardized inclusion and exclusion criteria. Consistent with systematic review protocol, articles were only removed by title and abstract if they clearly violated the eligibility criteria; others were moved forward to full text review. Inclusion of full text was decided by consensus by the 2 investigators, and when not reached, a third reviewer (K.L.) independently decided on inclusion.

\subsection*{Data Extraction and Outcomes}
Paired reviewers (J.F., A.H., K.L., S.B.) independently abstracted results from each article in duplicate. The included data from studies were: study characteristics, patient characteristics, type of colonoscopy performed (screening, surveillance, diagnostic), rate of bowel perforation, rate of severe GI bleeding, and length of follow up (Table 1). We expected that the available studies would be heterogeneous and the range of estimates quite wide; our intention was to identify a range of credible estimates for serious harm rather than summarizing the estimates in a formal meta-analysis. Credible estimates were drawn from the most robust studies using the McMaster Harms Tool, Table 2, which is the most widely used tool for evaluating the rigor and bias of research on harms.\textsuperscript{12}

The primary outcomes were serious harms – severe GI bleeding and GI perforation — secondary to screening colonoscopy. Consistent with 2 prior systematic reviews,\textsuperscript{8,9} severe bleeding was defined as requiring transfusion of at least 1 unit of packed red blood cells, admission to the hospital, postpolypectomy bleeding, or need for repeat endoscopic evaluation. Similarly, perforation was defined as free air or perforation visualized on radiograph requiring hospitalization or surgery.\textsuperscript{9} We contacted the authors of 2 eligible trials that did not report on specific duration of follow up or enrollment period.

\subsection*{Data Synthesis and Analysis}
A narrative synthesis of harm studies was completed. A meta-analysis of results was performed, but is not reported due to excess heterogeneity as anticipated, I\textsuperscript{2} >95%. When available, raw data were collected from
study data; when not available, the rate of harm per 10,000 individuals was calculated from data presented.

**Study Appraisal for Quality**

Quality was assessed by independent paired reviewers (S.B., J.F., A.H., K.L.) using the McMaster harms tool (Table 3). Despite being of variable quality, all studies meeting the selection criteria are reported on in this systematic review. The quality of each study is reported in Table 3. Due to low quality of some studies, the credible range of harms is presented from the most robust studies; specifically, these studies had the largest sample size, active method of collecting harms data, active follow up of harms, and included validation of the harms rate. Active methods for collecting harms included independent, individual review of charts and evaluation of claims data. Follow up of consequences of harms
included monitoring patients prospectively after colonoscopy. Finally, validation included comparison of patient reported harms, physician reported harms, and hospital administrative data.

**Results**

**Serious Harms**

Of 1951 citations identified, 1855 were excluded based on title and abstract alone with consensus by 2 independent reviewers (JF, AH). The kappa statistic representing inter-reviewer reliability was between 0.54 (2020 to 2022 review of 599 studies) - 0.75 (2002 to 2019 review of 1352 studies) and found to be moderate to substantial. Ninety-four studies were reviewed in full text, of which 6 met eligibility criteria and included only screening colonoscopies. Of those removed, twenty-five were removed due to a follow-up duration of less than 30 days; thirty-four were removed due to inadequate study design; sixteen were removed for including populations other than those at average risk; 7 were removed due to lack of report on outcomes; 4 were removed due to duplicate data sets, and 2 were removed due to nonresponse from the author. The rate of severe bleeding from all studies that met eligibility criteria ranged from 14.11 to 165.61 per 10,000 screening colonoscopies (Table 1). The rate of perforation from all eligible studies ranged from 0.08 to 9.15 per 10,000 screening colonoscopies (Table 1). Serious adverse events other than bleeding or perforation were not included in our analysis after extraction due to inconsistent reporting and methods of data collection.

**Credible Range of Harms**

Because of the large range of estimates from eligible studies, taking a median or mean approximation of the harms may not represent the true rate of harms. Consequently, we cite the range of rates rather than summarizing the estimates in a formal meta-analysis. The range of rates were abstracted from Rabeneck and Zwink for perforation and severe bleeding as these were the most robust studies using a best evidence approach (Table 2). The credible range of bleeding rates was 16.4 to 36.2 per 10,000 screening colonoscopies and the perforation rate range was 7.6 to 8.5 per 10,000.

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Year</th>
<th>Study Type</th>
<th>Mean Age/Minimum</th>
<th>Type of Colonoscopy</th>
<th>Percent Polypectomy (If Reported)</th>
<th>Severe Bleeding /10,000 Colonoscopies</th>
<th>Perforations /10,000 Colonoscopies</th>
<th>Cardiovascular Adverse Events /10,000 Colonoscopies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cardioma et al 44</td>
<td>2009</td>
<td>Prospective</td>
<td>&gt;55</td>
<td>Screening</td>
<td>Not reported</td>
<td>2.04</td>
<td>16.42</td>
<td>Not reported</td>
</tr>
<tr>
<td>Bretthauer 23</td>
<td>2016</td>
<td>Retrospective</td>
<td>60</td>
<td>Screening</td>
<td>48.10%</td>
<td>0.80</td>
<td>14.32</td>
<td>Not reported</td>
</tr>
<tr>
<td>Causado-Calo 26</td>
<td>2020</td>
<td>Retrospective</td>
<td>&gt;50</td>
<td>Screening</td>
<td>Not reported</td>
<td>5.0</td>
<td>3.56</td>
<td>Not reported</td>
</tr>
<tr>
<td>Rabeneck 25</td>
<td>2008</td>
<td>Retrospective</td>
<td>60.9</td>
<td>Screening</td>
<td>24.24%</td>
<td>3.56</td>
<td>14.11</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zwink 29</td>
<td>2014</td>
<td>Prospective</td>
<td>76.7</td>
<td>Screening</td>
<td>12.13%</td>
<td>10.04</td>
<td>113.0</td>
<td>Not reported</td>
</tr>
<tr>
<td>Zafar 26</td>
<td>2014</td>
<td>Retrospective</td>
<td>76.7</td>
<td>Screening</td>
<td>Not reported</td>
<td>12.13</td>
<td>Not reported</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion
Understanding and reducing harm is a common goal of clinicians and health systems.19 We conducted a rigorous analysis of harms that occurred in the course of screening colonoscopy, which is the third most common cancer screening procedure in the US, exceeded only by Papanicolaou tests and mammography.20 Recently, a publication in the New England Journal by Bretthauer et al. suggested that the mortality benefits of screening colonoscopy have been overestimated in cohort studies. A more precise estimate of harms of this screening tool was cited by the authors as relevant to health system and individual decision making.21 In the Bretthauer publication, bleeding was cited as a complication of screening colonoscopies, and occurred when polypectomies took place. Patients should consider the impact of irregular findings when considering screening mechanisms. Our study is unique in that it includes only studies that had at least 30 days of follow up, used similar definitions for significant harms, and provided data on colonoscopies completed for screening indications specifically.

Our estimates of harms are higher than those of the USPSTF, which cites the risk of severe bleeding from screening colonoscopy to be 2 in 10,000 procedures (95% CI, 0.7-4 in 10,000; $I^2 = 52.5\%$) and the risk of perforation 1 in 10,000 procedures (95% 0.4-1.4 in 10,000; $I^2 = 18.4$).3 This difference may be due to the shorter follow up period of studies included in their analysis: the USPSTF review included eligible studies regardless of length of follow up period. In the studies we reviewed, the vast majority of complications occurred after the first 72 hours, resulting in a higher harms estimate than that of the task force.

<table>
<thead>
<tr>
<th>Were there harms PRE-DEFINED using standardized or precise definitions?</th>
<th>Bokemeyer</th>
<th>Brethauer</th>
<th>Causada-Calo</th>
<th>Rabeneck</th>
<th>Zwink</th>
<th>Zafar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Were SERIOUS events precisely defined?</th>
<th>Yes</th>
<th>No</th>
<th>No</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Were SEVERE events precisely defined?</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Were the number of DEATHS in each study group specified OR were the reason(s) for not specifying them given?</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the mode of harms collection specified as ACTIVE?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Was the mode of harms collection specified as PASSIVE?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the study specify WHO collected the harms?</th>
<th>No</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the study specify the TRAINING or BACKGROUND of who ascertained the harms?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the study specify the TIMING and FREQUENCY of collection of the harms?</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the author(s) use STANDARD scale(s) or checklist(s) for harms collection?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the authors specify if the harms reported encompass ALL the events collected or a selected SAMPLE?</th>
<th>All</th>
<th>All</th>
<th>All</th>
<th>All</th>
<th>All</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the NUMBER of participants that withdrew or were lost to follow-up specified for each study group?</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td>Was the TOTAL NUMBER of participants affected by harms specified for each study arm?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the author(s) specify the NUMBER for each TYPE of harmful event for each study group?</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the author(s) specify the type of analyses undertaken for harms data?</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 2. Adapted McMaster Harms Evaluation with Best Practice Evaluation
A wide range of harms was found in the review of the literature. This may be due to the lack of standard measures for complications of colonoscopy and varied reporting methods. Not all complications of colonoscopy result in the same consequences for the patient, hospital, and health care system. A metric for analyzing harms that would assist further in identifying the true rate would include patient reported vs clinician identified harms; a scale of severity (such as degree of hypovolemia, severity of pain, or development of advanced disease); and whether intervention was required, such as Emergency Department utilization subsequent to the procedure; hospitalizations secondary to the procedure; and mortality. Although current electronic medical records (EMR) may limit the specific linkage between index procedure and complications, the EMR can be modified to create a more reliable record for following downstream complications of colonoscopy.

In a future analysis, this team will evaluate the impact of harms on the general population in the context of overuse of screening colonoscopy. It is likely that screening colonoscopy is overused, as cited by Djinoban, 2019 and in our own recent systematic review.17,24 Harms in the context of overuse place patients at unnecessary and potentially preventable risk.

### Strengths and Limitations

The observed rates of severe harms subsequent to screening colonoscopy varied widely, and studies appeared to suffer from systematic biases which would lead to an underestimate. For example, many of the studies captured in our systematic review used algorithms that search administrative data, which do not include instances of harm if they are not in some way linked to the index colonoscopy. Downstream harms, such as prolonged hospitalization and subsequent procedures, were seldom included in studies. Some studies relied on reports of harm by endoscopists. Other studies that compared patients’ clinical charts to endoscopists’ estimates of harms experienced by their patients found that the clinicians routinely provided underestimates.25 This study may also underestimate harms because of billing and EHR documentation of colonoscopies. If a screening colonoscopy results in a polypectomy, it may be documented as diagnostic; polypectomies are more closely related to serious harms. Thus,
we may underestimate due to the nomenclature of colonoscopies.

Our study may further underestimate the true rate of harm from colonoscopy because we included only severe perforations and GI bleeds (requiring hospitalization), whereas other types of severe harm were excluded. Multiple studies have identified a range of severe harms from colonoscopy, such as infections requiring hospital admission, cardiac arrest, stroke, heart attack, and death. But these harms could not be included in our study due to a scarcity of data. Numerous other rare but serious complications associated with colonoscopy have been described in case reports, but these harms are almost never reported in large population studies. Such rare but serious complications include splenic rupture, postpolypectomy syndrome, diverticulitis, mesenteric vessel rupture, and subcutaneous emphysema.

Given that we could not include these other types of severe harm, it is unlikely we have overestimated rates of severe harm and we most likely underestimated them.

Summary

Despite the limitations in the data on rates of harm related to screening colonoscopy, it is possible to estimate a credible range of rates. Further research into standardizing metrics and reporting for harms may reduce the size of this range, allowing better discussions between clinicians and patients on potential screening options. Patients should be actively engaged during decision making regarding colon cancer screening; the type of screening that a patient chooses should be based on their personal risk factors, goals, and values. In a future study, we will estimate the rate of harms incurred during overuse of colonoscopy.

We are grateful for the contributions of Robert Ballieu (methodology), Richelle Cooper (conceptualization), Julia Healey (correspondence), Jerry Hoffman (conceptualization), Anuradha Jetty (formal analysis), and Elizabeth Wilkinson (formal analysis).

To see this article online, please go to: http://jabfm.org/content/00/00/000.full.

References