Research Article

iMedPub Journals
www.imedpub.com

Journal of Clinical Gastroenterology and Hepatology
ISSN 2575-7733

2018
Vol.2 No.3:17

DOI: 10.21767/2575-7733.1000046

An Analysis of Fragility and Risk of Bias in Randomized Clinical Trials in Bowel Preparation Guidelines

Chris Chapman*, Benjamin Howard, Cole Wayant and Matt Vassar

Department of Institutional Research, Oklahoma State University Center for Health Sciences, USA

***Corresponding author**: Chris Chapman, Department of Institutional Research, Oklahoma State University Center for Health Sciences, USA, Tel: + 918-582-1972; E-mail: clobrynth918@aol.com

Rec date: June 25, 2018; **Acc date**: July 26, 2018; **Pub date**: July 30, 2018

Citation: Chapman C, Howard B, Wayan C, Vassar M (2018) An Analysis of Fragility and Risk of Bias in Randomized Clinical Trials in Bowel Preparation Guidelines. J Clin Gastroenterol Hepatol Vol.2: No.3:17.

Abstract

Objective: In this study, we use the Fragility Index and Cochrane's Risk of Bias Tool 2.0 to analyze the randomized controlled trials underpinning the American Gastroenterological Association's clinical practice guideline on bowel preparation before colonoscopy.

Design: All citations within the guideline were screened for specific criteria. We extracted bowel preparation outcome data from the included studies and used an online calculator to determine the FI and Fragility Quotient (FQ) (fragility index relative to study sample size). Risk of bias assessments was made using the Cochrane Risk of Bias Tool 2.0.

Results: The median FI for the 30 included trials was 7.5 events (IQR 3-11.75). The median FQ was 3.5 per 100 patients. The Risk of Bias Assessments resulted in the following classifications: 12: Low Risk, 2: Some Concerns, 16: High Risk.

Conclusion: RCTs in ACG Bowel Preparation guidelines were found to contain moderate fragility and relatively high risk of bias. Reporting fragility in RCTs will help appraisers of guidelines by indicating the robustness of the results. In this way, guideline writers will be in a better position to make recommendations. Likewise, pre-emptive evaluation of risk of bias will help identify key weaknesses underlying RCTs and add to their credibility in formulating recommendations.

Keywords: Clinical practice guidelines; Randomized controlled trials; Fragility index; Fragility quotient; Risk of bias; Bowel preparation; Colonoscopy; Colon cancer

Key Points

- What is already known about the subject?

The fragility index demonstrates the number of events that would need to change to alter the significance of a reported result. Several studies have already suggested baseline acceptable values for fragility index.

- What are the new findings?

Our investigation demonstrated that RCTs found in Bowel Preparation Guidelines have moderate fragility and high risk of bias, particularly in regard to selective reporting of outcomes.

- How might it impact clinical practice in the foreseeable future?

Reporting study fragility and preemptive assessment of bias would allow practitioners to assess the weight of results in RCTs and thus have more confidence in guideline recommendations.

Introduction

Clinical practice guidelines (CPGs) are essential to evidence-based clinical decision making. CPG recommendations are often based on a systematic survey of the literature and are graded on the strength of the recommendation and quality of evidence. Typically, the strongest recommendations are supported by methodologically robust evidence, like that from randomized controlled trials (RCTs). However, not all RCTs are equally robust. The strength of CPG recommendations depends on the robustness of the RCTs used as evidence.

In the field of gastroenterology, colonoscopy is a key step in the evaluation of colorectal carcinoma (CRC). CRC is the second most commonly diagnosed cancer in women and the third in men worldwide with an estimated 1.4 million new cases each year [1]. The incidence of CRC has declined in recent years within the US, likely due to increased screening and new technologies. To continue this trend, evidence-based guidelines with robust underlying evidence are needed.

The American Gastroenterological Association (AGA) guidelines for bowel cleansing for colonoscopy provide 30 evidence-based recommendations to endoscopists [2]. In accordance with the Grades of Recommendation Assessment, Development and Evaluation (GRADE) system, authors of CPGs and the GRADE system recognized RCTs as the highest level of evidence, along with systematic reviews [3]. Concerning, however, is the recent demonstration that the results of RCTs are often fragile [4,5]. The application of the fragility index (FI)

and fragility quotient (FQ) has shown that the statistical significance of clinical trial results often relies on relatively few patient events. To calculate the FI, one must iteratively remove one patient event from a study arm and add it to the other until the statistical significance of a clinical endpoint is nullified. The FQ contextualizes the FI by dividing the FI by the RCTs sample size. RCTs with fragile outcomes are less trustworthy, and the findings may be irreproducible.

Furthermore, bias may compromise the trustworthiness of RCT outcomes. To ensure trustworthy findings and improve transparency in RCTs, the Cochrane Risk of Bias tool 2.0 was developed [6]. The tool was designed to objectively analyze study elements to identify potential sources of bias. This tool assesses bias over five domains: randomization, deviation from intended interventions, accounting for missing data, measurements, and selectively reporting results. Each of the five domains requires equal attention to mitigate bias in RCTs.

Our primary objective in this study is to examine the fragility of the RCT endpoints that underpin the AGA guideline on bowel cleansing for colonoscopy using the FI and FQ. Our secondary objective was to assess the methodological quality of the RCTs using the Cochrane Risk of Bias Tool 2.0.

Methods

Identification of studies

Using the AGA CPGs, we identified all RCTs cited within the document. Using Pubmed and Google, we located the articles eligible for review. Two investigators (C.C., C.W.) screened, reviewed, and included all studies that met eligibility criteria.

Eligibility criteria

To be evaluated for fragility, studies must possess the following three characteristics: randomization between two patient groups in a 1:1 distribution, a parallel two group design, and at least one statistically significant dichotomous endpoint.

Data collection

Data was extracted from the articles using a Google form. Data collected includes the sample size for each group, number of patients lost to follow-up, the reported outcome, the rates of the reported outcome within the groups, statistical significance value, and the method for determining statistical significance. We prioritized primary outcomes, but included the secondary outcome if it met the criteria for our FI analysis and the primary outcome did not.

Fragility index and fragility index quotient

Fragility index was calculated from the extracted outcome data using an online calculator [7]. This calculator requires the number of events in each arm and the respective sample size. It first calculates a p-value based on the event rates using Fisher's exact test. If the p-value is statistically significant, the calculator iteratively subtracts and adds one patient event at a time until the p-value is non-significant. If the p-value is not statistically significant, the FI is 0. In such a case, the original statistically significant p-value was often obtained from a test other than Fisher's exact. The FQ was then calculated from the FI and total sample size **(Table 1)**.

Risk of bias assessment

We devised a Google Form based on the Cochrane Risk of Bias Tool 2.0 as a framework to assist in data collection. This tool examines the study design of randomized controlled trials and the way data is handled within those trials to determine if a potential for bias exists. Specific sources of bias are assessed over five domains: randomization, deviation from intended interventions, accounting for missing data, measurements, and selectively reporting results.

The tool provides a flowchart for each of the domains and one of three grades is given for each of the domains depending on the available information. The grades are as follows: low risk, some concerns, and high risk of bias. Two investigators (C.C., B.H.) reviewed articles independently before sharing notes. Conflicts were resolved by consensus. We assigned an overall grade for each of the RCTs after comparing the individual grades for the five domains. A study was considered to have a low risk of bias only if given low risk grades across all five domains. If the study has one domain that has some concern for bias but all other domains were low risk, then the study was classified as having some concern for bias. A final grade of high risk of bias was assigned to any study with 2 or more some concerns domains or any study with at least one high risk domain. Risk of bias assessments were limited to the 30 studies assessed for fragility.

Graphical representation of the data

Graphical representation of the data was done using R Studio software and the ggplot2 package [8,9].

Results

Study selection

The AGA's guideline on bowel preparation prior to colonoscopy contained 253 references. Of these, 133 were RCTs. Thirty met inclusion criteria and were included in the final analysis of fragility **(Figure 1)**. Of these, 27 (90.0%) primary outcomes and 3 (10.0%) secondary outcomes were recorded.

Overall fragility index and fragility quotient

The median FI for the 30 trials was 7.5 events (IQR 3 - 11.75). One study (1/30, 3.3%) had a FI of 0, indicating that the reported outcome was not significant according to Fisher's exact test. The median FQ for the trials was 0.035, meaning 3.5 per 100 patients (IQR 1.5 - 6.7 per 100) were needed to nullify the statistical significance of the RCT endpoints in our study. The median sample size was 200 (104.7 - 337.2). The number

lost to follow up was greater than or equal to the FI in 5 trials (5/30, 16.7%). Complete FI and FQ results can be found in **Appendix 1 (Figure 2)**.

Risk of bias assessment

Our analysis of the trials revealed 16 (53.3%) at high risk, 2 (6.7%) with some concerns for bias, and 12 (40.0%) at low risk of bias. High risk of bias most commonly came from selection in the reported result (30.0%) followed by deviations from expected interventions (20.0%). Cumulative risk of bias data is shown in **Table 2** with assessments for each RCT included in supplemental **Appendix 2**.

Discussion

Overall, our results show that RCTs in AGA Bowel Cleansing Guidelines have moderate fragility and a relatively high risk of bias, most often due to biased reporting of outcomes and bias arising from differences in care provided between treatment groups. Regarding fragility, the reported median FI across all 30 RCTs was 7.5 events. This implies that a median of only 7-8 events would be required to reverse the significance of a certain result. These results are consistent with previous fragility studies, which reported median fragility indices of 7 and 8 [4,5].

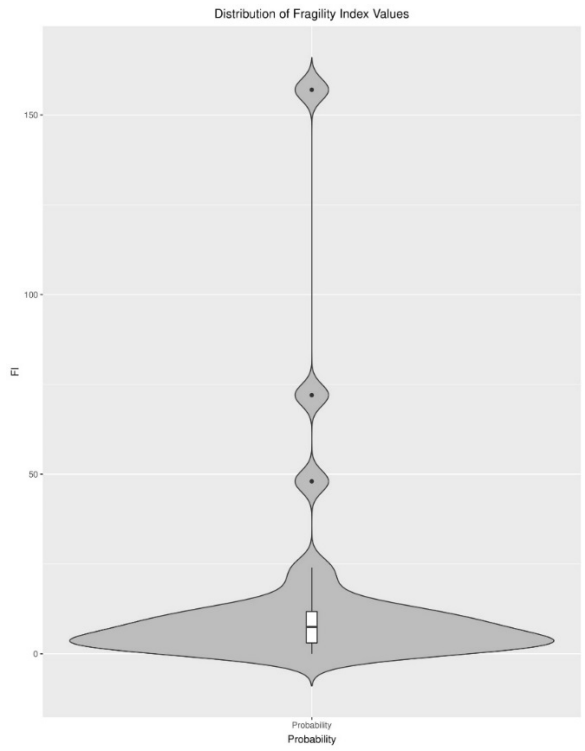

Figure 1 Prisma diagram of search strategy.

Table 1 Fragility assessment.

Randomized controlled trial	Outcome	Sample Size	Patients Lost to Follow Up	Intervention Group	Control Group	Outcome: Intervention	Outcome: Control	Fragility Index	Fragility Quotient
Abdul-Baki, et al. 2008	"Satisfactory" bowel cleanliness	382	0	199	183	177	78	72	0.188
Abut, et al. 2009	"Satisfactory" or "unsatisfactory" bowel cleanliness	80	4	41	39	41	23	12	0.15
Amato, et al. 2010	Presence of moderate to severe bowel pain	296	0	151	145	9	22	3	0.01
Arezzo,2000	"Good" or "Medium"bowel preparations	200	0	100	100	77	95	9	0.045
Berkelhammer, et al. 2002	"Good" or "Fair" Bowel Cleanliness	300	0	140	160	132	117	24	0.08
Cesaro, et al. 2013	"Excellent" Bowel Cleanliness	99	0	50	49	35	24	1	0.01
Cohen et al. 1994	"Excellent" bowel cleanliness	181	0	143	138	93	55	19	0.068
Cohen, et al. 2010	"Excellent" bowel preparation	107	3	52	55	36	21	7	0.065
Delegge, et al. 2005	"Satisfactory" bowel preparations	506	0	284	222	89	82	0	0
El- Baba, et al. 2006	"Excellent" bowel preparations	62	0	36	26	18	5	2	0.032

Haappama ki, et al. 2011	"Poor" or"inadeqaute" bowel cleanliness	399	85	203	196	22	8	3	0.008
Law, et al. 2004	"Excellent" or"good" bowel cleanliness	207	13	101	106	77	57	11	0.053
Lee, et al. 2010	Number of patients with an unsatisfactory bowel cleansing	104	0	51	53	23	42	8	0.077
Malik, et al. 2009	"Adequate" Bowel Cleanliness	81	0	40	41	39	31	3	0.037
Marmo, et al. 2010	"Satisfactory" Bowel Cleanliness	870	27	437	433	327	141	157	0.18
Park, et al. 2010	"Excellent" Bowel Cleanliness rated as excellent on Aronchik scale	152	38	73	79	55	40	8	0.053
Parra-Blanco, et al. 2006	Failed bowel preparations	177	20	89	88	11	2	2	0.011
Picchio, et al. 2008	"Adequate" Bowel cleanliness	182	0	94	88	67	46	5	0.027
Radaelli, et al. 2005	"Excellent" or "good"bowel cleansing	383	5	191	192	173	153	8	0.021
Repici, et al. 2012	Bowel Cleanliness with BPPP scores of 6 or greater.	377	28	187	190	148	133	1	0.003
Rex, et al. 2010	"Excellent" bowel preparations	130	7	63	67	45	23	14	0.108
Rex, et al. 2013	"Satisfactory" bowel cleanliness	601	2	304	297	256	221	11	0.018
Samarasena, et al.2012	"Excellent" bowel preparations	222	0	105	117	83	30	48	0.216
Saunders,et al. 1995	"Good" or"Excellent" bowelcleanliness	89	0	44	45	34	43	3	0.034
Sharara, et al. 2013	"Excellent" or better bowel cleanliness	99	0	49	50	31	17	5	0.051
Tae, et al. 2012	Patients receiving agood (>5) rating forbowel prep on BBPSscale.	200	5	102	98	95	80	3	0.015
Tajika, et al. 2012	Scores of excellent/ good/fair vs poor/ inadequate	244	0	119	125	39	24	3	0.012
Vradelis, et al. 2009	Adequate" colon cleansing	342	3	182	160	148	108	8	0.023
Young, et al. 2000	"Good" bowel cleanliness	323	0	169	154	144	105	13	0.04
Zwas, et al. 1996	Presence of aphthoid lesions	97	0	53	44	13	1	3	0.031

The FIs calculated in our investigation, in conjunction with FQs and risk of bias, raise questions about the robustness of the evidence underpinning the AGA's Bowel Cleansing Guidelines. Furthermore, 5 of the 30 (16.7%) trials reported a number lost to follow up greater than or equal to the fragility index. These lost participants may have been able to provide data that affected the occurrence of study outcomes thus altering the significance of reported results. Previous FI investigations have emphasized the effect that patients lost to follow up have on trial results and suggested that studies with FIs lower than the number of participants lost to follow up are more fragile [4,10]. Our findings highlight the utility of the FI and FQ by guideline panels who may wish to investigate the robustness of statistically significant endpoints. If guideline developers report FI and FQ values alongside p-values,

physician readers would be better equipped to make confident clinical decisions.

Figure 2 Distribution of fragility studies.

Table 2 General risk of bias assessments.

Risk of Bias Category (n=30)	Low Risk	Some Concerns	High Risk
Arising from the Randomization process	21 (70%)	8 (26.7%)	1 (3.3%)
Due to deviations from intended interventions	24 (80%)	0	6 (20%)
Due to missing outcome data	30 (100%)	0	0
Measurement of the Outcome	29 (96.7%)	0	1 (3.3%)
Selection of the reported Result	21 (70%)	0	9 (30%)
Overall bias of trial	12 (40%)	2 (6.7%)	16 (53.3%)

For guideline developers to implement and for physicians to understand the FI and FQ, they must understand the relationship between the two measures. For example, within our analysis, the FI ranged from 0 to 157. One may assume that the larger the FI, the more robust the trial outcome, since more patient events are necessary to nullify the statistical significance. However, differences in characteristics of each study affect the magnitude of FI. To contextualize the magnitude of the FI, one must calculate the FQ — the FI divided by the trial sample size [11]. Consider an example from our analysis; one RCT outcome had an FI of 157, while another had an FI of 48. But, the first trial had a FQ of 18 and the second had an FQ of 21. So, in fact, fewer patients per 100 were needed to nullify the statistical significance of the first trial, despite the magnitude of its FI.

Our results reveal several studies with low FI values, with one in particular having an FI of 0. The FI calculator begins by calculating a p-value using Fisher's exact test. For there to be an FI of zero, Fisher's exact test would have to yield a non-significant p-value. Therefore, the original p-value was significant due to choice of statistical test and likely irreproducible. For such an outcome to underpin CPG recommendations is concerning, again supporting the utility of the FI and FQ as standard measures for RCT authors and CPG developers. We assessed the FQ for the 5 studies with an FI less than 3 (the lower bound of the median FI IQR). All of these studies had an FQ less than or equal to 3 in 100 patients. In comparison, the studies with FI greater than the upper bound of the IQR had FQ values around 20 per 100 patients. While it is important to note that the studies with higher FQs require roughly 7x as many event changes to alter the significance of the reported result, few studies have reported FQ values. Therefore, normal FQ values across RCTs have not been clearly established.

Our investigation also revealed a high risk of bias in a large portion of RCTs. Following evaluation by the Cochrane Risk of Bias 2.0 Tool, 16 of the 30 RCTs (53.3%) were found to be at high risk of bias. High risk of bias was most frequently due to the selection in the reported result (30.0%) followed by deviations from expected interventions (20.0%). Selective reporting bias in research has been shown to contribute to misleading information and recommendations that are based off faulty evidence [12]. Preconceived notions or an interest in showing benefit of a particular intervention may lead researchers to selectively report on outcome measurements that are favorable to the intervention [6]. Our assessment of risk of bias emphasizes the importance of carefully examining RCTs for bias before clinical trial interventions are implemented into practice.

The main limitation to our study is that the results may not be generalizable to all 133 RCTs from the guideline. The inclusion criteria for the fragility arm allowed evaluation of only 30 of 133 RCTs. Due to the nature of the fragility analysis, this limitation was unavoidable. Despite this limitation, our results show that analyzing studies for fragility and ROB will help authors provide the best recommendations from the available data.

Conclusion

Making clinical decisions affects patients and as such it is crucial that practitioners have confidence in the recommendations in clinical practice guidelines. It is understood that readers should not rely on any one particular value, however we advocate for an exhaustive approach when evaluating RCTs and their contribution to CPGs. Inclusion of

the FI, FQ, and ROB for RCTs that provide the evidence for CPGs would improve the confidence in the data and recommendations. Based on our investigation, it is clear that RCTs in the ACG Bowel Preparation guidelines exhibit moderate fragility and a relatively high risk of bias and this should be taken into account when implementing these recommendations into clinical practice.

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, et al. (2015) Global cancer statistics, 2012. CA: CA Cancer J Clin 65: 87-108.
2. Johnson DA, Barkun AN, Cohen LB, Dominitz JA, Kaltenbach T, et al. (2014) Optimizing adequacy of bowel cleansing for colonoscopy: recommendations from the US multi-society task force on colorectal cancer. Am J Gastroenterol 109: 1528.
3. http://training.cochrane.org/path/grade-approach-evaluating-quality-evidence-pathway.
4. Matics TJ, Khan N, Jani P, Kane JM (2017) The fragility index in a cohort of pediatric randomized controlled trials. J Clin Med 6: 79.
5. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, et al. (2014) The statistical significance of randomized controlled trial results is frequently fragile: A case for a fragility index. J Clin Epidemiol 67: 622-8.
6. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, et al. (2011) The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Bmj 343: d5928.
7. Kane SP (2017) Fragility Index Calculator. 2017. http://clincalc.com/Stats/FragilityIndex.aspx.
8. R-Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
9. Wickham H (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, USA.
10. Mazzinari G, Ball L, Neto AS, Errando CL, Dondorp AM, et al. (2018) The fragility of statistically significant findings in randomised controlled anaesthesiology trials: A systematic search of the medical literature. Br J Anaesth.
11. Ahmed W, Fowler RA, McCredie VA (2016) Does sample size matter when interpreting the fragility index? Crit. Care Med 44: 1142-3.
12. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, et al. (2010) The impact of outcome reporting bias in randomized controlled trials on a cohort of systematic reviews. BMJ 340: 365.