Trust but Verify: Our experiences identifying fabricated or falsified trials in a systematic review

Dena Zeraatkar, PhD

Assistant Professor, McMaster University zeraatd@mcmaster.ca





Conflicts of Interest

Member of the GRADE Working Group

Member of Cochrane

Consultancies

- European Headache Federation
- Society for Evidence-Based Gender Medicine (SEGM)
- Association of Medical Microbiology and Infectious Disease Canada



Scientific Integrity



Scientific Integrity

Fabricated Study Data	Falsified Study Data	Studies with errors in execution or analysis

Tools to Detect and Manage Problematic Studies

Reference	ΤοοΙ	Domains/Description
Moussa 2024	Research Integrity in Guidelines and evIDence synthesis (RIGID)	A framework describing the integration of integrity assessments in evidence synthesis and guideline development
Mol 2023	TRACT Checklist	Governance, author group, plausibility of intervention, timeframe, dropouts, baseline characteristics, and outcomes
Weibel 2022	Research Integrity Assessment Tool	Retraction or expression of concern, trial registration, ethics approval, author group, methods, results
Alfirevic 2021	Cochrane Pregnancy and Childbirth Trustworthiness Screening Tool (CPC-TST)	Research governance, baseline characteristics, feasibility
Grey 2020	REAPPRAISED checklist	Research governance, ethics, authorship, productivity, plagiarism, research conduct, analyses and methods, image manipulation, statistics and data, errors, data duplication and reporting

Problematic Trials



Baseline Characteristics Improbable with Randomization

Table 1. Sociodemographic and Clinical Characteristics of the Patients at Baseline.*

The NEW ENGLAND JOURNAL of MEDICINE

ABLISHED IN 1812 NOVEMBER 12, 2020 VOL. 383 NO. 20

A Randomized Trial Comparing Antibiotics with Appendectomy for Appendicitis

ABSTRACT	
AACCORDUND Antibiotic therapy has been proposed as an alternative to surgery for the treatment of appendicitis.	The mem (D.R. Flun N.I. Shapi F.T. Drake
We conducted a pragmatic, nonlinficiently, randomized trial compar- ge authoric theory. (10) deap course with appendecome in patients with append- tish at 25 U.S. centers. The primary outcome was 30-day health status, as assessed with the European Quilty of Life-5 bitmensions (EQ-50) exceptionalite focores range from 0 to 1, with higher scores indicating better health status, nonlinefrio- ry margin, 0.05 points. Secondary outcome included appendications in the slow patient of the status of the status of the status of the status slow patient of the status of the status of the slow patient of the slow	Patton, H batini, B.A. V. Sohn, I Chung, D. Chiang, A Foster, S. Mandell, I DeUgarte, man, H.B. Thompsor busch, R.J dasan, E.
total, 1552 adults (414 with an appendicolith) underwent randomization: 776	Kessler, ar

ty margin, 0.05 points). Secondary outcomes included appendectomy in the anti- biotics group and complications through 90 days; analyses were prespecified in subgroups defined according to the presence or absence of an appendicolith.	DeUgarte, A.H. Kaji, G.J. Moran, D. Saltz- man, H.B. Alam, P.K. Park, L.S. Kao, C.M. Thompson, W.H. Self, J.T. Yu, A. Wie- busch, R.J. Winchell, S. Clark, A. Krishna- dasan, E. Fannon, D.C. Lavallee, B.A.
terest. In coal, 152 adults (414 with an appendicolith) underwent randomization; 776 were assigned to receive antibiotics (67% of whom were not hospitalized for the disc treatment and and 756 to undergo approximations to approximately of the approximation of the stream of the stream of the stream of the approximation of the stream of	Comstock, B. Bizzell, RJ, Heagerty, LG, Kessler, and DA. Talan) assume respon- sibility for the overall content and integ- ring of this attice. The fail names, aca- demic degrees, and affiliations of the members of the writing committee are listed in the Appendix. Address reprint requests to Dr. Fham at the Surgical Out- comiss Research Center, Department of Surgery, University of Washington, Box 356400, Seattle, WA 98395-6410, or at dwelfumgluwedu.
group than in the appendectomy group (8.1 vs. 3.5 per 100 participants; rate ratio, 2.28; 95% CI, 1.30 to 3.98); the higher rate in the antibiotics group could be attrib- ted to those with an appendicolith (20.2 vs. 3.6 per 100 participants; rate ratio,	*A complete list of members of the CODA Collaborative is provided in the Supple- mentary Appendix, available at NEJM.org.
5.69; 95% CI, 2.11 to 15.38) and not to those without an appendicolith (3.7 vs. 3.5 per 100 participants; rate ratio, 1.05; 95% CI, 0.45 to 2.43). The rate of serious	This article was published on October 5, 2020, at NEJM.org.
adverse events was 4.0 per 100 participants in the antibiotics group and 3.0 per 100 participants in the appendectomy group (rate ratio, 1.29; 95% CI, 0.67 to 2.50).	N Engl J Med 2020;383:1907-19. DOI: 10.1056/NEJMox2014320 Copyright © 2020 Messechasetts Medical Society.

CONCLUSIONS

For the rearment of appendicits, antibiotics were noninferior to appendecomy on the basis of results of a standard beath-straus measure. In the antibiotics group, nearly 3 in 10 participants bad undergone appendecomy by 90 days. Participants with an appendicibility were at a higher risk for appendecomy and for complications than those without an appendicolith. (Funded by the Patiento-Centered Oucomes Research Institute, COOM CollicalPital, Systa Number, NCIU2800785)

> N ENGL J MED 348320 NEJM.ORG NOVEMBER 12, 2020 The New England Journal of Medicine Downloaded from nejm org on September 1, 2024. For personal use only,

Characteristic	Antibiotics (N=776)	Appendectomy (N = 776)	
Age — yr	38.3±13.4	37.8±13.7	P value
Sex — no. (%)			
Female	286 (37)	290 (37)	
Male	490 (63)	486 (63)	P value
Gender different from sex assigned at birth — no. (%)	8 (1)	6 (1)	P I TUIN
Race or ethnic group — no. (%)†			
White	461 (60)	449 (59)	
Black	75 (10)	63 (8)	
American Indian or Alaska Native	13 (2)	9 (1)	→ P value
Asian	39 (5)	53 (7)	
Native Hawaiian or Pacific Islander	4 (1)	3 (<1)	
Multiple or other	176 (23)	185 (24)	
Hispanic ethnic group†	362 (47)	366 (47)	→ P value
Primary language — no. (%)			
English	469 (60)	464 (60)	
Spanish	267 (34)	267 (34)	- P value
Other	40 (5)	45 (6)	F value
insurance — no. (%)			
Commercial	323 (43)	317 (42)	
Medicare or Tricare	89 (12)	89 (12)	- P valu
Medicaid or other state program	134 (18)	131 (17)	F Value
Other or no coverage	213 (28)	217 (29)	
Modified Charlson comorbidity index score:	0.24±0.53	0.24±0.53	P value
Body-mass index§	29.0±6.6	28.6±6.1	P value
Duration of symptoms — days	1.8±3.6	1.6±1.6	P value
Alvarado score¶	6.6±1.6	6.7±1.7	P value
History of fever — no. (%)	194 (25)	185 (24)	P value
initial white-cell count — per µl	12,900±4000	13,400±4100	
Imaging test — no. (%)		14.9	- i valu
Computed tomography alone	626 (81)	609 (78)	
Ultrasonography alone	24 (3)	30 (4)	P value
>1 Imaging test	125 (16)	137 (18)	



Prevalence of Issues Related to Research Integrity

Citation	Sample	Methods	Estimate
		The number of retractions in the scientific literature has	4462 (45/101)
Weeks 2023	374 randomized trials in Cochrane reviews of pregnancy and childbirth	been estimated at less than 1%, a figure which leads some to	• 1/374)
		believe that research integrity concerns are not a widespread	8 0/35)
Carlisle 2020	526 randomized trials submitted to Anesthesiology	statistical summaries and tests, interrogation of IPD (repetition and duplication, end digit preference),	8 I; 44 an ng those wi ac ss to IPD
			3 tals

Prevalence of Issues Related to Research Integrity

Citation	Sample	Methods	Estimate
Mousa 2024	101 randomized trials considered for gidelines addressing PCOS	TRACT checklist	44.6% (45/101)
Weeks 2023	374 randomized trials in Cochrane reviews of pregnancy and childbirth	Cochrane Pregnancy and Childbirth Trustworthiness Screening Tool (CPC-TST)	24.9% (93/374)
Bordewijk 2020	35 randomized trials inwomen's health published by2 authors	Identical or similar values in baseline characteristics; compatibility of baseline characteristics with chance	85.7% (30/35)
Carlisle 2020	526 randomized trials submitted to Anesthesiology	Probability of baseline characteristics, plausibility of statistical summaries and tests, interrogation of IPD (repetition and duplication, end digit preference),	8% overall; 44% among those with access to IPD
Roberts 2007	Trials investigating mannitol for head injury	Investigation by Cochrane Collaboration + several other institutions	3 trials

Living Systematic Review of Interventions for the Management of Long COVID

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental rehabilitation program vs	161 more per 1,000 (61 more to 292 more)	-2 (-3.96 to -0.04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1 (-1.98 to -0.02)	0.04 (0 to 0.08)	20 more per 1,000
Usual care	RR: 1.55 (1.21 to 2)a				-1.5 (-2.41 to -0.59)		(10 fewer to 50 more)
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1,000 (30 fewer to 30 more)
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	200 more per 1,000 (94 more to 336 more) RR: 1.47 (1.22 to 1.79)a Concentration 239 more per 1,000 (112 more to 401 more) RR: 1.52 (1.29 to 2.04)b Dyspina 150 more per 1,000 (27 more to 250 more) RR: 1.28 (1.05 to 154)c					1.5 (-0.87 to 3.87)	0 more per 1,000 (10 fewer to 10 more)
Intermittent aerobic exercise vs. Continuous aerobic exercise			3.8 (1.12 to 6.48)		0 (-3.69 to 3.69)		
Transcranial direct current stimulation, Physiotherapy, Education related to	315 more per 1,000 (59 more to 699 more)						0 more per 1,000

Living Systematic Review of Interventions for the Management of Long COVID

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental	161 more per 1,000 (61 more to 292 more)	-2 (-3 96 to -0 04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1 (-1.98 to -0.02)	0.04 (0 to 0.08)	20 more per 1,000
Usual care	RR: 1.55 (1.21 to 2)a	2(3.50 10 0.04)	0.5 (1.01 (0 1.01)	1 (0.44 (0 2.44)	-1.5 (-2.41 to -0.59)	0.04 (0 10 0.00)	(10 fewer to 50 more)
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1,000 (30 fewer to 30 more)
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	Fatigue 200 more per 1,000 (94 more to 356 more) RR: 147 (1.22 to 1.79)a Concentration 239 more per 1,000 (112 more to 401 more) RR: 152 (1.29 to 2.04)b Dypones 150 more per 1,000 (27 more to 290 more) RR: 128 (1.05 to 154)c					15 (-0.87 to 3.87)	0 more per 1,000 (10 fewer to 10 more)
Intermittent aerobic exercise vs. Continuous aerobic exercise			3.8 (1.12 to 6.48)		0 (-3.69 to 3.69)		
Transcranial direct current stimulation, Physiotherapy, Education related to	315 more per 1,000 (59 more to 699 more)						0 more per 1,000

In response to growing concerns about untrustworthy trial publications, we incorporated methods to assess trials for signs of fabrication, falsification, or major errors.





Search

Screening

Data collection

Analysis

Certainty of evidence

Search

Screening

Data collection



MEDLINE EMBASE CINAHL PsycInfo AMED CENTRAL EPISTEMONIKOS COVID-19 repository Inception to December 2023

Analysis

Certainty of evidence



Search Screening Data collection Analysis



Eligibility criteria

- Adults (≥18 years old)
- Long COVID: symptoms at three or more months following laboratory confirmed, probable, or suspected COVID-19 infection that persist for at least two months
- Randomized to any pharmacologic or nonpharmacologic intervention(s), placebo, sham, usual care
- Min 25 patients/arm

Certainty of evidence



Ρ

Search

Screening

Data collection

Analysis

Certainty of evidence

Data	RoB 2	
rial characteristics Country Registration Design atient characteristics Age, Sex Diagnostic criteria Time since infection Duration of long COVID symptoms Comorbidities	Outcomes of interest:-Recovery or improvement-Fatigue-Post-exertional malaise-Patient-reported function-Cognitive function-Mental health-Dyspnea-Quality of life-Changes in education/ employment status-Serious adverse events	TRACT CHECKLIST Governance Author group Plausibility of intervention Timeframe Dropouts Baseline characteristics Outcomes



evidence







evidence



Search

Screening

Data collection

Analysis

Certainty of evidence

Data (RoB 2	
Trial characteristics - Country - Registration - Design Patient characteristics - Age, Sex - Diagnostic criteria - Time since infection	Outcomes of interest: - Recovery or improvement - Fatigue - Post-exertional malaise - Patient-reported function - Cognitive function - Mental health - Dyspnea	TRACT CHECKLIS Governance Author group
 Duration of long COVID symptoms Comorbidities 	 Quality of life Changes in education/ employment status Serious adverse events 	 Plausibility of intervention Timeframe Dropouts Baseline characteristi Outcomes

Search

Screening

Data collection

Analysis

Certainty of evidence

Data Co	RoB 2	
Frial characteristics Country Registration Design Patient characteristics Age, Sex Diagnostic criteria Time since infection Duration of long COVID symptoms	Outcomes of interest: - Recovery or improvement - Fatigue - Post-exertional malaise - Patient-reported function - Cognitive function - Mental health - Dyspnea - Quality of life - Changes in education/	TRACT CHECKLIST Governance Author group Plausibility of intervention Timeframe Dropouts
Comorbidities	- Serious adverse events	Baseline characteristic Outcomes

TRACT Checklist

Governance	Author group	Plausibility of intervention	Timeframe	Dropouts	Baseline characteristics	Outcomes
Absent or retrospective registration	Three or fewer authors	Insufficient or implausible description	Implausibly short time between ending recruitment/foll ow up and submission of the paper	Zero patients lost to follow- up despite long follow-up period	Perfect balance for multiple baseline characteristics	Effect size that is much larger than in other RCTs regarding the same topic



more domains

Characteristics of Problematic Trials



One third of trials contained issues that raised concerns about their integrity and were ultimately disregarded from our review.

Concerns about integrity



Characteristics of Problematic Trials

registration

single center



Number of trials Concerns Fewer than three authors Author with history of retraction(s) Not registered Retrospective registration Critical design differences between trial report and trial Inconceivably fast recruitment of participants within a Improbably small number of participants (or 0 participants) discontinued the trial Baseline characteristics unlikely with randomization Suspicious outcome data Implausibly positive results

Example Problematic Trial



Abstract

Introduction: Post-pandemic syndrome has lasting functional and psychological consequences, especially for the elderly. This timeline requires a quick search for procedures that will enable us to implement safe and non-invasive therapeutic instruments as prophylactic or adjuvant therapies for post-COVID-19 consequences. Photobiomodulation (PBM) may decrease inflammation and improve leg circulation. So, this study aims to assess the impact of PBM on post-CO-VID-19 functional capacity and fatigability.

Material and methods: Two groups of 100 elders with a positive COVID-19 history were established. The PBM gro-

- I. Even, round numbers
- 2. Equal numbers of participants randomized to each arm without block randomization
- 3. Remarkably similar baseline characteristics across arms
- 4. 0 attrition
- 5. Exceptionally small variability in outcome measures
- 6. Trial registration describes a different trial
- 7. Author with a history of retractions due to research integrity issues

Comparison	/important	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious advers events
	improvement			Effect estimates			
	161 more per 1,000 (61 more to 292 more)				-1 (-1.98 to -0.02)		20 more per 1 000
rehabilitation program vs. Usual care	RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1.5 (-2.41 to -0.59)	0.04 (0 to 0.08)	(10 fewer to 50 mor
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1,000 (30 fewer to 30 mor
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	Fatigue 200 more per 1.000 (94 more to 336 more) Rit: 1.47 (1.22 to 1.79)a Concentration 239 more per 1.000 (112 more to 401 more) Rit: 1.52 (1.29 to 2.04)b Dysprea 150 more per 1.000 (27 more to 290 more) Rit: 1.28 (1.05 to 1.54)c					1.5 (-0.87 to 3.87)	0 more per 1,000 (10 fewer to 10 mor
Intermittent aerobic exercise vs. Continuous aerobic exercise			3.8 (1.12 to 6.48)		0 (-3.69 to 3.69)		
Transcranial direct current stimulation, Physiotherapy, Education related to activities of daily living vs. Physiotherapy, Education related to self-management	315 more per 1,000 (59 more to 699 more) RR: 1.69 (1.13 to 2.53)d	-12.4 (-17.33 to -7.47)			-4.91 (-7.5 to -2.32)	14.8 (8.86 to 20.74)	0 more per 1,000 (50 fewer to 50 mor
Multicomponent exercise of progressively increasing intensity, Physiotherapy vs. Physiotherapy			6.96 (2.7 to 11.22)		2.06 (-3.52 to 7.64)		
Hyperbaric oxygen therapy vs. Usual care	ĺ		-5.2 (-14.06 to 3.66)	3.4 (0.3 to 6.5)	-7.1 (-12.23 to -1.97) 10 (-0.01 to 20.01)		
Vortioxetine vs. Usual care				-0.02 (-0.24 to 0.2)	-1.59 (-3 to -0.18)	2.36 (0.71 to 4.01)	
Telerehabilitation app ('ReCOVery') vs. Usual Care			-3.46 (-9.07 to 2.15)	0.61 (-0.9 to 2.12)	1.87 (-5.39 to 9.13)		0 more per 1,000 (40 fewer to 40 mor
Leronlimab vs. Usual care Inspiratory muscle training vs. Usual care		-0.08 (-0.65 to 0.49)		0.08 (-0.45 to 0.61)	0.03 (-0.45 to 0.51)	-1.3 (-5.9 to 3.3)	
Amygdala and insula retraining vs. Education related to self-management		-1.48 (-3 to 0.04)					
Coenzyme Q10 vs. Usual Care						-0.04 (-0.1 to 0.02)	0 more per 1,000 (30 fewer to 30 mor
L-arginine, vitamin C vs. Usual care	826 more per 1,000 (155 more to 3366 more) RR: 10.5 (2.78 to						0 more per 1,000 (80 fewer to 80 mor
Glucosaminyl muramyl			6.88 (2.92 to 10.84)		-2.47 (-4.52 to -0.42)		

High	Definitely	Definitely	Definitely no				
certainty	more effective	worse	different				
Moderate	Probably	Probably	Probably no				
certainty	more effective	worse	different				
Low	May be more	May be worse	May be no				
certainty	effective		different				
Very low certainty	We are very uncertain						

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adv events
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1 (-1.98 to -0.02) -1.5 (-2.41 to -0.59)	0.04 (0 to 0.08)	20 more per 1, (10 fewer to 50 r
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1, (30 fewer to 30 r
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	Fatigue 200 more per 1,000 (94 more to 336 more) RR: 1.47 (1.22 to 1.79)a Concentration 239 more per 1,000 (112 more to 401 more) RR: 1.62 (1.29 to 2.04)b Dyspnea 150 more per 1,000 (27 more to 290 more)					1.5 (-0.87 to 3.87)	0 more per 1,0 (10 fewer to 10 r
Intermittent aerobic	RR: 1.28 (1.05 to 1.54)c		2.0 (1.42) - 5.40)		0 (2 (24+ 2 (2))		
aerobic exercise			3.8 (1.12 (0 0.48)		0 (-3.09 (0 3.09)		
Transcranial direct current stimulation, Physiotherapy, Education related to activities of daily living vs. Physiotherapy, Education related to self-management	315 more per 1,000 (59 more to 699 more) RR: 1.69 (1.13 to 2.53)d	-12.4 (-17.33 to -7.47)			-4.91 (-7.5 to -2.32)	14.8 (8.86 to 20.74)	0 more per 1,((50 fewer to 50 r
Multicomponent exercise o progressively increasing intensity, Physiotherapy vs Physiotherapy	F		6.96 (2.7 to 11.22)		2.06 (-3.52 to 7.64)		
Hyperbaric oxygen therapy			-5.2 (-14.06 to 3.66)	3.4 (0.3 to 6.5)	-7.1 (-12.23 to -1.97)		
vs. Usual care				-0.02 (-0.24 to 0.2)	10 (-0.01 to 20.01) -1.59 (-3 to -0.18)	2.36 (0.71 to 4.01)	
Telerehabilitation app ('ReCOVery') vs. Usual Care			-3.46 (-9.07 to 2.15)	0.61 (-0.9 to 2.12)	1.87 (-5.39 to 9.13)		0 more per 1,0 (40 fewer to 40
Leronlimab vs. Usual care		-0.08 (-0.65 to 0.49)		0.08 (-0.45 to 0.61)	0.03 (-0.45 to 0.51)		
Inspiratory muscle training vs. Usual care						-1.3 (-5.9 to 3.3)	
Amygdala and insula retraining vs. Education related to self-management		-1.48 (-3 to 0.04)					
Coenzyme Q10 vs. Usual Care						-0.04 (-0.1 to 0.02)	0 more per 1,0 (30 fewer to 30
L-arginine, vitamin C vs.	826 more per 1,000 (155 more to 3366 more)						0 more per 1,0
							(80 fewer to 80 r

High	Definitely	Definitely	Definitely no					
certainty	more effective	worse	different					
Moderate	Probably	Probably	Probably no					
certainty	more effective	worse	different					
Low	May be more	May be worse	May be no					
certainty	effective		different					
Very low certainty	We are very uncertain							

Comparisons

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious advers events
				Effect estimates			
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1 (-1.98 to -0.02) -1.5 (-2.41 to -0.59)	0.04 (0 to 0.08)	20 more per 1,000 (10 fewer to 50 mor
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1,000 (30 fewer to 30 mo
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	Fatigue 200 more per 1,000 (94 more to 336 more) Concentration 239 more per 1,000 (112 more to 401 more) RR: 15.6 (1.29 to 2.04)b Dypone 150 more per 1,000 (27 more to 290 more) RR: 1.28 (1.05 to 1.54)c					15 (-0.87 to 3.87)	0 more per 1,000 (10 fewer to 10 mo
Intermittent aerobic exercise vs. Continuous aerobic exercise			3.8 (1.12 to 6.48)		0 (-3.69 to 3.69)		
Transcranial direct current stimulation, Physiotherapy, Education related to activities of daily living vs. Physiotherapy, Education related to self-management	315 more per 1,000 (59 more to 699 more) RR: 1.69 (1.13 to 2.53)d	-12.4 (-17.33 to -7.47)			-4.91 (-7.5 to -2.32)	14.8 (8.86 to 20.74)	0 more per 1,001 (50 fewer to 50 mc
Multicomponent exercise of progressively increasing intensity, Physiotherapy vs. Physiotherapy			6.96 (2.7 to 11.22)		2.06 (-3.52 to 7.64)		
Hyperbaric oxygen therapy			-5.2 (-14.06 to 3.66)	3.4 (0.3 to 6.5)	-7.1 (-12.23 to -1.97)		
vs. Usual care					10 (-0.01 to 20.01)		
Vortioxetine vs. Usual care				-0.02 (-0.24 to 0.2)	-1.59 (-3 to -0.18)	2.36 (0.71 to 4.01)	
Telerehabilitation app ('ReCOVery') vs. Usual Care			-3.46 (-9.07 to 2.15)	0.61 (-0.9 to 2.12)	1.87 (-5.39 to 9.13)		0 more per 1,000 (40 fewer to 40 mo
		-0.08 (-0.65 to 0.49)		0.08 (-0.45 to 0.61)	0.03 (-0.45 to 0.51)		
Inspiratory muscle training vs. Usual care						-1.3 (-5.9 to 3.3)	
Amygdala and insula retraining vs. Education related to self-management		-1.48 (-3 to 0.04)					
Coenzyme Q10 vs. Usual Care						-0.04 (-0.1 to 0.02)	0 more per 1,000 (30 fewer to 30 mo
L-arginine, vitamin C vs. Usual care	826 more per 1,000 (155 more to 3366 more)						0 more per 1,000 (80 fewer to 80 mo
Glucosaminyl muramyl	R: 10.5 (2.78 10		C 00 /2 02 to 40 5 *		-2.47 (-4.52 to -0.42)		
dinonsido ("Linonid") vo			6.88 (2.92 to 10.84)				

High	Definitely	Definitely	Definitely no					
certainty	more effective	worse	different					
Moderate	Probably	Probably	Probably no					
certainty	more effective	worse	different					
Low	May be more	May be worse	May be no					
certainty	effective		different					
Very low certainty	We are very uncertain							

Comparison	/important	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04)	0.5 (-1.01 to 2.01)	1 (-0.44 to 2.44)	-1 (-1.98 to -0.02) -1.5 (-2.41 to -0.59)	0.04 (0 to 0.08)	20 more per 1,000 (10 fewer to 50 more
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69)	4.9 (-1.89 to 11.69)	-5.2 (-7.97 to -2.43)			0 more per 1,000 (30 fewer to 30 more
A combinationn of probiotics and prebiotics ('Synbiotics') called SIM01 vs. Usual care	Fatigue 200 more per 1.000 (94 more to 336 more) RR: 1.47 (1.22 to 1.79)a Concentration 238 more per 1.000 (112 more to 401 more) RR: 1.52 (1.29 to 2.04)b Dyspnea 150 more per 1.000 (27 more to 230 more) RR: 1.28 (1.05 to 1.54)c					1.5 (-0.87 to 3.87)	0 more per 1.000 (10 fewer to 10 more
Intermittent aerobic exercise vs. Continuous aerobic exercise			3.8 (1.12 to 6.48)		0 (-3.69 to 3.69)		
Transcranial direct current stimulation, Physiotherapy, Education related to activities of daily living vs. Physiotherapy, Education related to self-management	315 more per 1,000 (59 more to 699 more) RR: 1.69 (1.13 to 2.53)d	-12.4 (-17.33 to -7.47)			-4.91 (-7.5 to -2.32)	14.8 (8.86 to 20.74)	0 more per 1,000 (50 fewer to 50 more
Multicomponent exercise of progressively increasing intensity, Physiotherapy vs. Physiotherapy			6.96 (2.7 to 11.22)		2.06 (-3.52 to 7.64)		
Hyperbaric oxygen therapy vs. Usual care			-5.2 (-14.06 to 3.66)	3.4 (0.3 to 6.5)	-7.1 (-12.23 to -1.97) 10 (-0.01 to 20.01)	2 36 (0 71 to 4 01)	
Vortioxetine vs. Usual care Telerehabilitation app ('ReCOVery') vs. Usual Care			-3.46 (-9.07 to 2.15)	0.61 (-0.9 to 2.12)	1.87 (-5.39 to 9.13)		0 more per 1,000 (40 fewer to 40 more
Leronlimab vs. Usual care Inspiratory muscle training		-0.08 (-0.65 to 0.49)		0.08 (-0.45 to 0.61)	0.03 (-0.45 to 0.51)	-1.3 (-5.9 to 3.3)	
vs. Usual care Amygdala and insula retraining vs. Education related to self-management		-1.48 (-3 to 0.04)					
Coenzyme Q10 vs. Usual Care	936 mars par 1 000					-0.04 (-0.1 to 0.02)	0 more per 1,000 (30 fewer to 30 more
L-arginine, vitamin C vs. Usual care	(155 more to 3366 more)						0 more per 1,000 (80 fewer to 80 more
Glucosaminyl muramyl	RR: 10.5 (2.78 to		6.88 (2.92 to 10.84)		-2.47 (-4.52 to -0.42)		

High	Definitely	Definitely	Definitely no					
certainty	more effective	worse	different					
Moderate	Probably	Probably	Probably no					
certainty	more effective	worse	different					
Low	May be more	May be worse	May be no					
certainty	effective		different					
Very low	We are very uncertain							

GRADE ratings and interpretation

High	Definitely	Definitely	Definitely no				
certainty	more effective	worse	different				
Moderate	Probably	Probably	Probably no				
certainty	more effective	worse	different				
Low	May be more	May be worse	May be no				
certainty	effective		different				
Very low certainty	We are very uncertain						

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04) PROMIS (patient- reported outcomes measurement information system)- Fatigue subscore (Mean: 50, SD: 10; higher scores indicate greater impairment)	0.5 (-1.01 to 2.01) PROMIS (patient- reported outcomes measurement information system)- Physical function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	1 (-0.44 to 2.44) PROMIS (patient- reported outcomes measurement information system)- Cognitive function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	-1 (-1.98 to -0.02) Hospital Anxiety and Depression Scale (HADS) anxiety subscale (Range: 0 to 21; higher scores indicate greater impairment) -1.5 (-2.41 to -0.59) Hospital Anxiety and Depression Scale (HADS) depression subscale (Range: 0 to 21; higher scores indicate greater impairment)	0.04 (0 to 0.08) PROMIS 29+2 Profile v2.1 (PROPr) (HRQoL) (Range: -0.022 to 1; higher scores indicate less impairment)	20 more per 1,000 (10 fewer to 50 more)
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69) Checklist Individual Strength (CIS) fatigue subscale (Range: 8 to 56; higher scores indicate greater impairment)	4.9 (-1.89 to 11.69) SF-36 Physical function subscale (Range: 0 to 100; higher scores indicate less impairment)	-5.2 (-7.97 to -2.43) Checklist Individual Strength (CIS) concentration problems subscale (Range: 4 to 28; higher scores indicate greater impairment)			0 more per 1,000 (30 fewer to 30 more)

GRADE ratings and interpretation

High	Definitely	Definitely	Definitely no				
certainty	more effective	worse	different				
Moderate	Probably	Probably	Probably no				
certainty	more effective	worse	different				
Low	May be more	May be worse	May be no				
certainty	effective		different				
Very low certainty	We are very uncertain						

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04) PROMIS (patient- reported outcomes measurement information system)- Fatigue subscore (Mean: 50, SD: 10; higher scores indicate greater impairment)	0.5 (-1.01 to 2.01) PROMIS (patient- reported outcomes measurement information system)- Physical function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	1 (-0.44 to 2.44) PROMIS (patient- reported outcomes measurement information system)- Cognitive function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	 -1 (-1.98 to -0.02) Hospital Anxiety and Depression Scale (HADS) anxiety subscal (Range: 0 to 21; highe scores indicate greate impairment) -1.5 (-2.41 to -0.59) Hospital Anxiety and Depression Scale (HADS) depression subscale (Range: 0 to 21; higher scores indicate greater 	0.04 (0 to 0.08) PROMIS 29+2 Profile v2.1 (PROPr) (HRQoL) (Range: -0.022 to 1; higher scores indicate less impairment)	20 more per 1,000 (10 fewer to 50 more)
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69) Checklist Individual Strength (CIS) fatigue subscale (Range: 8 to 56; higher scores indicate greater impairment)	4.9 (-1.89 to 11.69) SF-36 Physical function subscale (Range: 0 to 100; higher scores indicate less impairment)	-5.2 (-7.97 to -2.43) Checklist Individual Strength (CIS) concentration problems subscale (Range: 4 to 28; higher scores indicate greater impairment)			0 more per 1,000 (30 fewer to 30 more)

GRADE ratings and interpretation

High	Definitely	Definitely	Definitely no	
certainty	more effective	worse	different	
Moderate	Probably	Probably	Probably no	
certainty	more effective	worse	different	
Low	May be more	May be worse	May be no	
certainty	effective		different	
Very low certainty	We	tain		

Comparison	Recovery /important improvement	Fatigue	Physical function	Cognitive function	Mental health	Quality of life / Wellbeing	Serious adverse events
				Effect estimates			
Physical and mental rehabilitation program vs. Usual care	161 more per 1,000 (61 more to 292 more) RR: 1.55 (1.21 to 2)a	-2 (-3.96 to -0.04) PROMIS (patient- reported outcomes measurement information system)- Fatigue subscore (Mean: 50, SD: 10; higher scores indicate greater impairment)	0.5 (-1.01 to 2.01) PROMIS (patient- reported outcomes measurement information system)- Physical function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	1 (-0.44 to 2.44) PROMIS (patient- reported outcomes measurement information system)- Cognitive function abilities subscore (Mean: 50, SD: 10; higher scores indicate less impairment)	-1 (-1.98 to -0.02) Hospital Anxiety and Depression Scale (HADS) anxiety subscale (Range: 0 to 21; higher scores indicate greater impairment) -1.5 (-2.41 to -0.59) Hospital Anxiety and Depression Scale (HADS) depression subscale (Range: 0 to 21; higher scores indicate greater impairment)	0.04 (0 to 0.08) PROMIS 29+2 Profile v2.1 (PROPr) (HRQoL) (Range: -0.022 to 1; higher scores indicate less impairment)	20 more per 1,000 (10 fewer to 50 more)
CBT vs. Usual care	326 more per 1,000 (100 more to 695 more) RR: 2.24 (1.38 to 3.64)b	-8.4 (-13.11 to -3.69) Checklist Individual Strength (CIS) fatigue subscale (Range: 8 to 56; higher scores indicate greater impairment)	4.9 (-1.89 to 11.69) SF-36 Physical functio subscale (Range: 0 to 100; higher scores indicate less impairment)	-5.2 (-7.97 to -2.43) Checklist Individual Strength (CIS) concentration problem: subscale (Range: 4 to 28; higher scores indicate greater impairment)			0 more per 1,000 (30 fewer to 30 more)

Pharmacologic	Vortioxetine, leronlimab, glucosaminyl muramyl dipeptide ('Licopid'), actovegin
Physical activity and rehabilitation	Acupuncture, inspiratory muscle training, active cycle of breathing
Behavioral interventions	Mobile educational application ('Recovery'), amygdala and insula retraining
Diet and dietary supplements	Co-enzyme Q10, L-arginine and liposomal vitamin C, combination of trimethyl hydrazinium propionate and ethyl methyl hydroxy pyridine succinate ('Brainmax')
Medical devices and technologies	Hyperbaric oxygen, active high-definition transcranial direct current stimulation, photobiomodulation, active hydrogen therapy
Interventions for anosmia/ hyposmia	Alpha-lipoic acid, mometasone furoate nasal spray, a combination of ultramicronized palmitoylethanolamide and luteolin, pentasodium diethylenetriamine pentaacetate intranasal spray, injections of cerebrolysin

Pharmacologic	Vortioxetine, leronlimab, glucosaminyl muramyl dipeptide ('Licopid'), actovegin
Physical activity and rehabilitation	Acupuncture, inspiratory muscle training, active cycle of breathing
Behavioral interventions	Mobile educational application ('Recovery'), amygdala and insula retraining
Diet and dietary supplements	Co-enzyme Q10, L-arginine and liposomal vitamin C, combination of trimethyl hydrazinium propionate and ethyl methyl hydroxy pyridine succinate ('Brainmax')
Medical devices and technologies	Hyperbaric oxygen, active high-definition transcranial direct current stimulation, photobiomodulation , active hydrogen therapy
Interventions for anosmia/ hyposmia	Alpha-lipoic acid, mometasone furoate nasal spray, a combination of ultramicronized palmitoylethanolamide and luteolin, pentasodium diethylenetriamine pentaacetate intranasal spray, injections of cerebrolysin

Lessons Learned

Our experience performing this exercise suggests that problematic trials are very common and so reviewers should certainly be vigilant and incorporate research integrity checks.

Lessons Learned



It's difficult to judge whether a trial is fabricated or whether data have been falsified

Lessons Learned

COVID-19 trial preprints and published reports: Trustworthiness and impact

Dena Zeraatkar, PhD Assistant Professor |Department of Anesthesia | Department of Health Research Methods, Evidence, and Impact McMaster University <u>zeraatd@mcmaster.ca</u> **@**@denazera



zeraatd@mcmaster.ca



Lawrence Mbuagbaw Thomas Agoritsas Signe A. Flottorp Paul Garner Tyler Pitre Rachel Couban Jason W. Busse

