

REPRODUCIBILITY IN THERAPEUTIC RESEARCH

Florian Naudet, MD, PhD



COIs

None in the past 3 years

Fundings



Has Cochrane lost its way?

Dissent over growing centralisation culminated in the expulsion of one of Cochrane's founding members. **Melanie Newman** reports on the organisation's internal struggles

Melanie Newman *freelance journalist, London, UK*

Has Cochrane lost its way?

Dissent over growing centralisation culminated in the expulsion of one of Cochrane's founding members. **Melanie Newman** reports on the organisation's internal struggles

Melanie Newman *freelance journalist, London, UK*

The more radical fringes of Cochrane campaigned for access to raw trial data and clinical study reports, looking for evidence beyond industry funded trials and analysis.

Has Cochrane lost its way?

Dissent over growing centralisation culminated in the expulsion of one of Cochrane's founding members. **Melanie Newman** reports on the organisation's internal struggles

Melanie Newman *freelance journalist, London, UK*

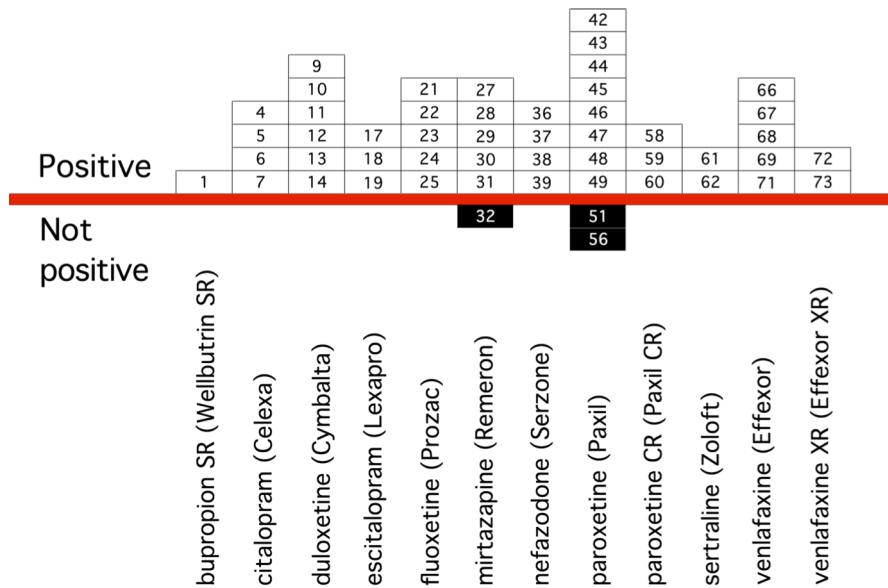
The more radical fringes of Cochrane campaigned for access to raw trial data and clinical study reports, looking for evidence beyond industry funded trials and analysis.

RAPID RESPONSE :

« **Is it "radical" and "fringe" to want access to raw data and to clinical study reports when we know how often negative data are left unpublished and when we know that positive outcomes often evaporate when unpublished data are included ?** »

03 January 2019

Jeanne Lenzer, journalist, independent and BMJ associate editor
New York



Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S., Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D.



			9		21	26		42				
		4	10	17	22	27		43				
		11	18	23	28	36		44		66		
		12	19	24	29	37		45		67	72	
Positive	1	5	12	19	24	30	38	48	58	61	69	73
	2	6	13	20	25	31	39	49	60	62	70	74
Not	3	7	14			32	40	50		63	71	
positive		8	15			33	41	51		64		
			16			34		52		65		
						35		53				
								54				
								55				
								56				
								57				
	bupropion SR (Wellbutrin SR)											
	citalopram (Celexa)											
	duloxetine (Cymbalta)											
	escitalopram (Lexapro)											
	fluoxetine (Prozac)											
	mirtazapine (Remeron)											
	nefazodone (Serzone)											
	paroxetine (Paxil)											
	paroxetine CR (Paxil CR)											
	sertraline (Zoloft)											
	venlafaxine (Effexor)											
	venlafaxine XR (Effexor XR)											

Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S.,
Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D.

Methods reproducibility – enough detail available to enable a study to be repeated;

Results reproducibility – the findings are replicated by others;

Inferential reproducibility – similar conclusions are drawn about results, which brings statistics and interpretation squarely into the mix.

SCIENTIFIC INTEGRITY

What does research reproducibility mean?

Steven N. Goodman,* Daniele Fanelli, John P. A. Ioannidis

Methods reproducibility – enough detail available to enable a study to be repeated;

Results reproducibility – the findings are replicated by others;

Inferential reproducibility – similar conclusions are drawn about results, which brings statistics and interpretation squarely into the mix.

SCIENTIFIC INTEGRITY

What does research reproducibility mean?

Steven N. Goodman,* Daniele Fanelli, John P. A. Ioannidis

Objective: To compare paroxetine with placebo and imipramine with placebo for the treatment of adolescent depression. **Method:** After a 7- to 14-day screening period, 275 adolescents with major depression began 8 weeks of double-blind paroxetine (20–40 mg), imipramine (gradual upward titration to 200–300 mg), or placebo. The two primary outcome measures were endpoint response (Hamilton Rating Scale for Depression [HAM-D] score ≤ 8 or $\geq 50\%$ reduction in baseline HAM-D) and change from baseline HAM-D score. Other depression-related variables were (1) HAM-D depressed mood item; (2) depression item of the Schedule for Affective Disorders and Schizophrenia for Adolescents-Lifetime version (K-SADS-L); (3) Clinical Global Impression (CGI) improvement scores of 1 or 2; (4) nine-item depression subscale of K-SADS-L; and (5) mean CGI improvement scores. **Results:** Paroxetine demonstrated significantly greater improvement compared with placebo in HAM-D total score ≤ 8 , HAM-D depressed mood item, K-SADS-L depressed mood item, and CGI score of 1 or 2. The response to imipramine was not significantly different from placebo for any measure. Neither paroxetine nor imipramine differed significantly from placebo on parent- or self-rating measures. Withdrawal rates for adverse effects were 9.7% and 6.9% for paroxetine and placebo, respectively. Of 31.5% of subjects stopping imipramine therapy because of adverse effects, nearly one third did so because of adverse cardiovascular effects. **Conclusions:** Paroxetine is generally well tolerated and effective for major depression in adolescents. *J. Am. Acad. Child Adolesc. Psychiatry*, 2001, 40(7):762–772. **Key Words:** paroxetine, imipramine, major depression, adolescent.

Efficacy of Paroxetine in the Treatment of Adolescent Major Depression: A Randomized, Controlled Trial

MARTIN B. KELLER, M.D., NEAL D. RYAN, M.D., MICHAEL STROBER, Ph.D., RACHEL G. KLEIN, Ph.D.,
STAN P. KUTCHER, M.D., BORIS BIRMAHER, M.D., OWEN R. HAGINO, M.D., HAROLD KOPLIEWICZ, M.D.,
GABRIELLE A. CARLSON, M.D., GREGORY N. CLARKE, Ph.D., GRAHAM J. EMSLIE, M.D.,
DAVID FEINBERG, M.D., BARBARA GELLER, M.D., VIVEK KUSUMAKAR, M.D.,
GEORGE PAPTAEODOROU, M.D., WILLIAM H. SACK, M.D., MICHAEL SWEENEY, Ph.D.,
KAREN DINEEN WAGNER, M.D., Ph.D., ELIZABETH B. WELLER, M.D., NANCY C. WINTERS, M.D.,
ROSEMARY OAKES, M.S., AND JAMES P. McCAFFERTY, B.S.

CONCLUSIONS

Neither paroxetine nor high dose imipramine showed efficacy for major depression in adolescents, and there was an increase in harms with both drugs. Access to primary data from trials has important implications for both clinical practice and research, including that published conclusions about efficacy and safety should not be read as authoritative. The reanalysis of Study 329 illustrates the necessity of making primary trial data and protocols available to increase the rigour of the evidence base.

Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence

Joanna Le Noury,¹ John M Nardo,² David Healy,¹ Jon Jureidini,³ Melissa Raven,³ Catalin Tufanaru,⁴ Elia Abi-Jaoude⁵

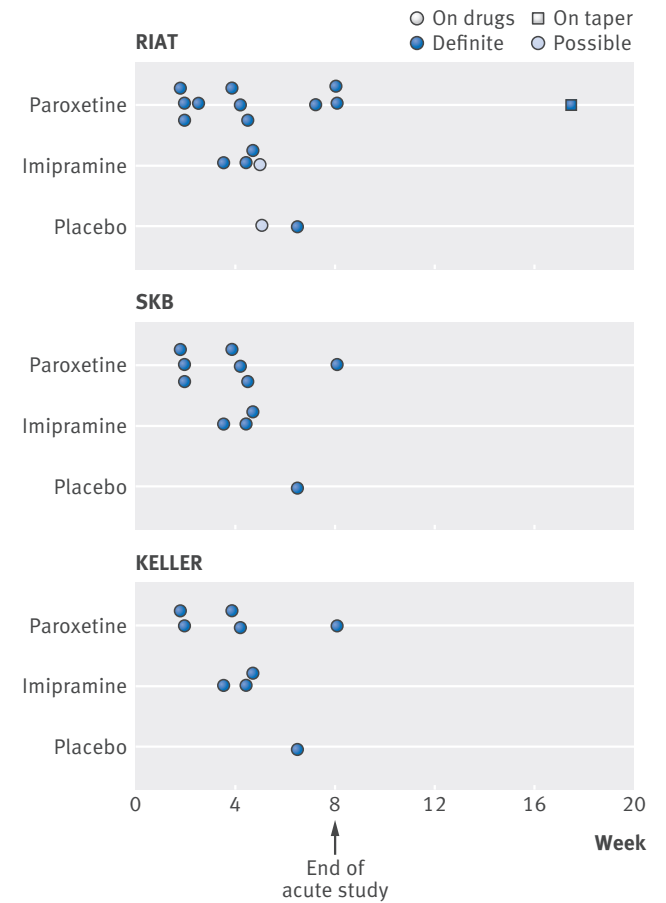


Fig 4 | Timing of suicidal and self injurious events in Study 329, Keller and colleagues, and RIAT analysis

IMPORTANCE Reanalyses of randomized clinical trial (RCT) data may help the scientific community assess the validity of reported trial results.

OBJECTIVES To identify published reanalyses of RCT data, to characterize methodological and other differences between the original trial and reanalysis, to evaluate the independence of authors performing the reanalyses, and to assess whether the reanalysis changed interpretations from the original article about the types or numbers of patients who should be treated.

DESIGN We completed an electronic search of MEDLINE from inception to March 9, 2014, to identify all published studies that completed a reanalysis of individual patient data from previously published RCTs addressing the same hypothesis as the original RCT. Four data extractors independently screened articles and extracted data.

MAIN OUTCOMES AND MEASURES Changes in direction and magnitude of treatment effect, statistical significance, and interpretation about the types or numbers of patients who should be treated.

RESULTS We identified 37 eligible reanalyses in 36 published articles, 5 of which were performed by entirely independent authors (2 based on publicly available data and 2 on data that were provided on request; data availability was unclear for 1). Reanalyses differed most commonly in statistical or analytical approaches ($n = 18$) and in definitions or measurements of the outcome of interest ($n = 12$). Four reanalyses changed the direction and 2 changed the magnitude of treatment effect, whereas 4 led to changes in statistical significance of findings. Thirteen reanalyses (35%) led to interpretations different from that of the original article, 3 (8%) showing that different patients should be treated; 1 (3%), that fewer patients should be treated; and 9 (24%), that more patients should be treated.

CONCLUSIONS AND RELEVANCE A small number of reanalyses of RCTs have been published to date. Only a few were conducted by entirely independent authors. Thirty-five percent of published reanalyses led to changes in findings that implied conclusions different from those of the original article about the types and number of patients who should be treated.

JAMA. 2014;312(10):1024-1032. doi:10.1001/jama.2014.9646

Original Investigation

Reanalyses of Randomized Clinical Trial Data

Shanil Ebrahim, PhD; Zahra N. Sohani, MSc; Luis Montoya, DDS; Arnav Agarwal, BSc; Kristian Thorlund, PhD; Edward J. Mills, PhD; John P. A. Ioannidis, MD, DSc

Annals of Internal Medicine

EDITORIAL

Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors

The International Committee of Medical Journal Editors (ICMJE) believes that there is **an ethical obligation** to responsibly share data generated by interventional clinical trials because **participants have put themselves at risk**.

In a growing consensus, many funders around the world—foundations, government agencies, and industry—now mandate data sharing. Here we outline ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

Annals of Internal Medicine

EDITORIAL

Sharing Clinical Trial Data: A Proposal From the International Committee of Medical Journal Editors

Annals of Internal Medicine

EDITORIAL

Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors

Methods

Study Population

The data, analytic methods, and study materials will not be made available to other researchers for purposes of reproducing the results or replicating the procedure. Study data are from NHANES (National Health and Nutrition Examination Survey) 2003–2006 cycles. NHANES samples

Moderate-to-Vigorous Physical Activity and All-Cause Mortality: Do Bouts Matter?

Pedro F. Saint-Maurice, PhD; Richard P. Troiano, PhD; Charles E. Matthews, PhD; William E. Kraus, MD

Data sharing and re-analysis for randomised controlled trials in leading biomedical journals with a full data-sharing policy: a survey of studies published in The BMJ and PLOS Medicine

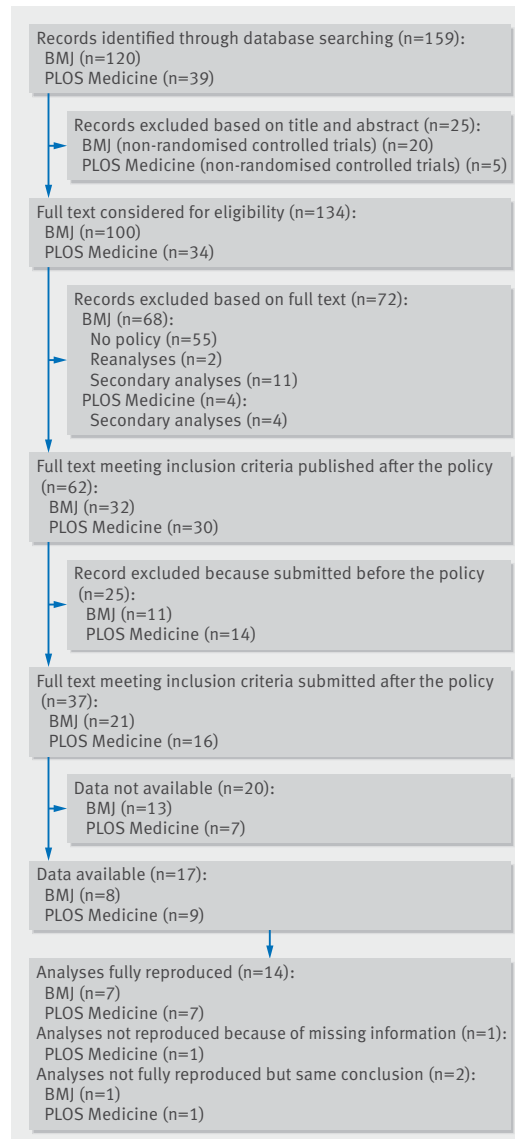
Florian Naudet, Charlotte Sakarovitch, Perrine Janiaud, Ioana Cristea, Daniele Fanelli, David Moher, John P.A. Ioannidis

The logo for 'thebmj' is displayed in white text on a blue rounded rectangular background.

We surveyed RCTs published in these 2 journals and to explore data availability and to perform re-analyses of the primary outcomes.



osf.io/u6hcv



Data availability: 17/37 (46%)

Analyses fully reproduced: 14/17 (82%)

Of the 3 remaining RCTs, **errors were identified in two but reached similar conclusions** and **one paper did not provide enough information in the Methods section** to reproduce the analyses.

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Geographical area of the lead country			
Europe	25 (67 %)	17 (80 %)	8 (50 %)
Australia and New Zealand	4 (11 %)	1 (5 %)	3 (19 %)
Northern America	3 (8 %)	1 (5 %)	2 (12.5 %)
Africa	3 (8 %)	1 (5 %)	2 (12.5 %)
East Asia	1 (3 %)	0 (0 %)	1 (6 %)
Middle East	1 (3 %)	1 (5 %)	0 (0 %)
Type of intervention			
Drug	20 (54 %)	13 (62 %)	7 (44 %)
Device	8 (22 %)	8 (38 %)	0 (0 %)
Complex intervention	9 (24 %)	0 (0 %)	9 (56 %)
Medical speciality			
Infectious disease	12 (33 %)	4 (19 %)	8 (50 %)
Rheumatology	5 (14 %)	5 (24 %)	0 (0 %)
Endocrinology/nutrition	4 (11 %)	1 (5 %)	3 (19 %)
Paediatrics	3 (8 %)	2 (9 %)	1 (6 %)
Mental health / addiction	2 (5 %)	1 (5 %)	1 (6 %)
Obstetrics	2 (5 %)	1 (5 %)	1 (6 %)
Emergency medicine	2 (5 %)	2 (9 %)	0 (0 %)
Geriatrics	2 (5 %)	0 (0 %)	2 (13 %)
Other	5 (14 %)	5 (24 %)	0 (0 %)

Table 1: Characteristics of the included studies. Numbers (and percentages) are presented (rounded percentages add up to 100% for each variable). For sample size medians and interquartile ranges are presented.

	All (37 studies)	BMJ (21 studies)	PLOS Medicine (16 studies)
Designs			
Superiority (Head to head)	18 (49 %)	15 (71 %)	3 (19 %)
Superiority (Factorial)	1 (3 %)	1 (5 %)	0 (0 %)
Superiority (Clusters)	8 (21 %)	1 (5 %)	7 (43 %)
Non-inferiority + Superiority (Head to head)	4 (11 %)	1 (5 %)	3 (19 %)
Non-inferiority (Head to head)	6 (16 %)	3 (14 %)	3 (19 %)
Sample size	432 (213 – 1070)	221 (159 – 494)	1047 (433 – 2248)
Private sponsorship			
No	26 (70 %)	15 (71 %)	11 (69 %)
Provided the device	1 (3 %)	1 (5 %)	0 (0 %)
Provided the intervention	1 (3 %)	0 (0 %)	1 (6 %)
Provided the drug	5 (13 %)	1 (5 %)	4 (25 %)
Provided the drug and some financial support	2 (5 %)	2 (9 %)	0 (0 %)
Provided partial financial support	1 (3 %)	1 (5 %)	0 (0 %)
Provided total financial support	1 (3 %)	1 (5 %)	0 (0 %)

Table 1: Characteristics of the included studies. Numbers (and percentages) are presented (rounded percentages add up to 100% for each variable). For sample size medians and interquartile ranges are presented.

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)

...

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)

Dear Florian Naudet,

please find attached the data-spreadsheet from our trial « XXX »

The variables are all labelled in a way that should be self-explanatory, if you require further explanation, I am very happy to answer you questions and support your work.

Best regards,

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)



Data custodian

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)



Data custodian

APPLICATION FOR RESEARCH DATA

This application form **MUST** be completed for all data requests. Protocols may sent in support of applications but all sections of this form must be fully completed [do not simply write “refer to protocol”]

Summary information

Brief study title	Data sharing and re-analysis for randomised controlled trials in leading biomedical journals with a full data-sharing policy: a survey of studies published in the BMJ and PLOS Medicine
Lead applicant name	FLORIAN NAUDET
Date of application	02/14/2017
Source trial title	Comparison of the two most commonly used treatments for pyoderma gangrenosum: results of the STOP GAP randomised controlled trial.
Reference number	This will be assigned by NCTU

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)



Data custodian



Data Sharing and Use Agreement

This agreement governs the terms on which access will be granted to the trial data detailed below. In signing this agreement the data requester is agreeing to be bound by the terms and conditions of access set out in this agreement. The terms of access set out in this agreement apply both to the data requester and the data requester’s Institution.

Title of the requesting study	Data sharing and re-analysis for randomised controlled trials in leading biomedical journals with a full data-sharing policy: a survey of studies published in the BMJ and PLOS Medicine
Title of the study being shared	Comparison of the two most commonly used treatments for pyoderma gangrenosum: results of the STOP GAP randomised controlled trial
Application reference	07371601

Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)



Data from: A novel brief therapy for patients who attempt suicide: a 24-months follow-up randomized controlled study of the Attempted Suicide Short Intervention Program (ASSIP)

Gysin-Maillart A, Schwab S, Soravia LM, Megert M, Michel K

Date Published: March 3, 2016

DOI: <https://doi.org/10.5061/dryad.85nf3>



Table 1 | Characteristics of included studies. Values are numbers (percentages) unless stated otherwise

Characteristics	All (37 studies)	<i>The BMJ</i> (21 studies)	<i>PLOS Medicine</i> (16 studies)
Statement of availability:			
Ask to contact by email	23 (62)	17 (81)	6 (38)
Explain how to retrieve data (eg, platform)	9 (24)	0 (0)	9 (56)
State “no additional data available”	2 (5)	2 (9)	0 (0)
Ask to contact by mail	1 (3)	0 (0)	1 (6)
Embargo	1 (3)	1 (5)	0 (0)
No statement	1 (3)	1 (5)	0 (0)

[LOGIN](#)



[SHARE](#) [f](#) [t](#)

[Data Catalog](#) [Citations](#) [Policies and Procedures](#) [Acknowledgements](#) [Contacts](#)

[HOME](#) > [CENTRAL DATA CATALOG](#) > [IMPACT-EVALUATION](#) > [DD_IHI_HEALTH_INSIST_201310_V00](#)

Southern Tanzania - Improving newborn survival in Southern Tanzania endline impact survey 2013: Mtwara and Lindi regions

	Reference ID	DD_IHI_HEALTH_INSIST_201310_V00	CREATED ON	Feb 05, 2015
	Year	0	LAST MODIFIED	Feb 09, 2015
	Country	Southern Tanzania	PAGE VIEWS	3615
	Producer(s)	Joanna Schellenberg - London School of Hygiene and Tropical Medicine, London, UK		
	Sponsor(s)	Save the Children - -		
	Collection(s)	Health Systems, Impact Evaluation and Policy		
Metadata	Documentation in PDF			

Perceived costs/benefit ratio of re-analyses

Costs involved in the data-sharing process

“As the study was launched we did not plan the cost for this preparation”

“Took some time to translate it [...] the original one was done in Hebrew”

“[We] decided to do the work for free although this is some extra work. For future projects it will be important to consider these costs either on your [side] or in the grant application for the trials”

Perceived benefits of sharing data for the purpose of this study

“We could create such a dataset [...] but it would require substantial effort and we cannot do it simply to demonstrate that it is possible.”

“We are especially keen that our data are used for IPD meta-analyses and have shared this with [...] we see that as an exemplar of meaningful data-sharing. Yours is a most unusual request”

“A slight concern about ‘naming and shaming’ individual studies/investigators”

Novelty and heterogeneity in data-sharing practices

Some authors who were unsure how to proceed

"[...] However, I am just wanting to confirm School policy and our ethical obligations regarding the sharing of data before we proceed"

"Please can you let me know how you have been receiving data from other centers securely?"

Heterogeneity between different procedures to share data

Open repository (n=5)

Downloadable on a secured website (n=1) after registration

Included as appendix of the published paper (n=3)

Sent by e-mail (n=10).

In 3 occasions, we signed a data-sharing request/agreement. In addition, typically there was no standard in type of data-shared. In one case, authors mentioned explicitly that they followed standardized guidelines¹⁵ to prepare the dataset.

Incomplete or ambiguous labels and reporting

Complexity of some analyses

Obtaining more information about the analytic method by contacting authors was sometimes (6 studies) necessary

Incomplete information

Three databases did not provide sufficient information to reproduce the analyses:

- Variables used for adjustment
- Definition of the analysis population
- Randomization groups

Communication with authors was therefore necessary and was fruitful in one these 3 cases.

Of course...

2 very selected journals (selected studies)

Reproducibility of their analysis VS the best standards



What's next ?

Reproducibility in therapeutic research



2 PhD students

A twitter account



@ReiTheR_RCT



A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group’s data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as “research parasites.”

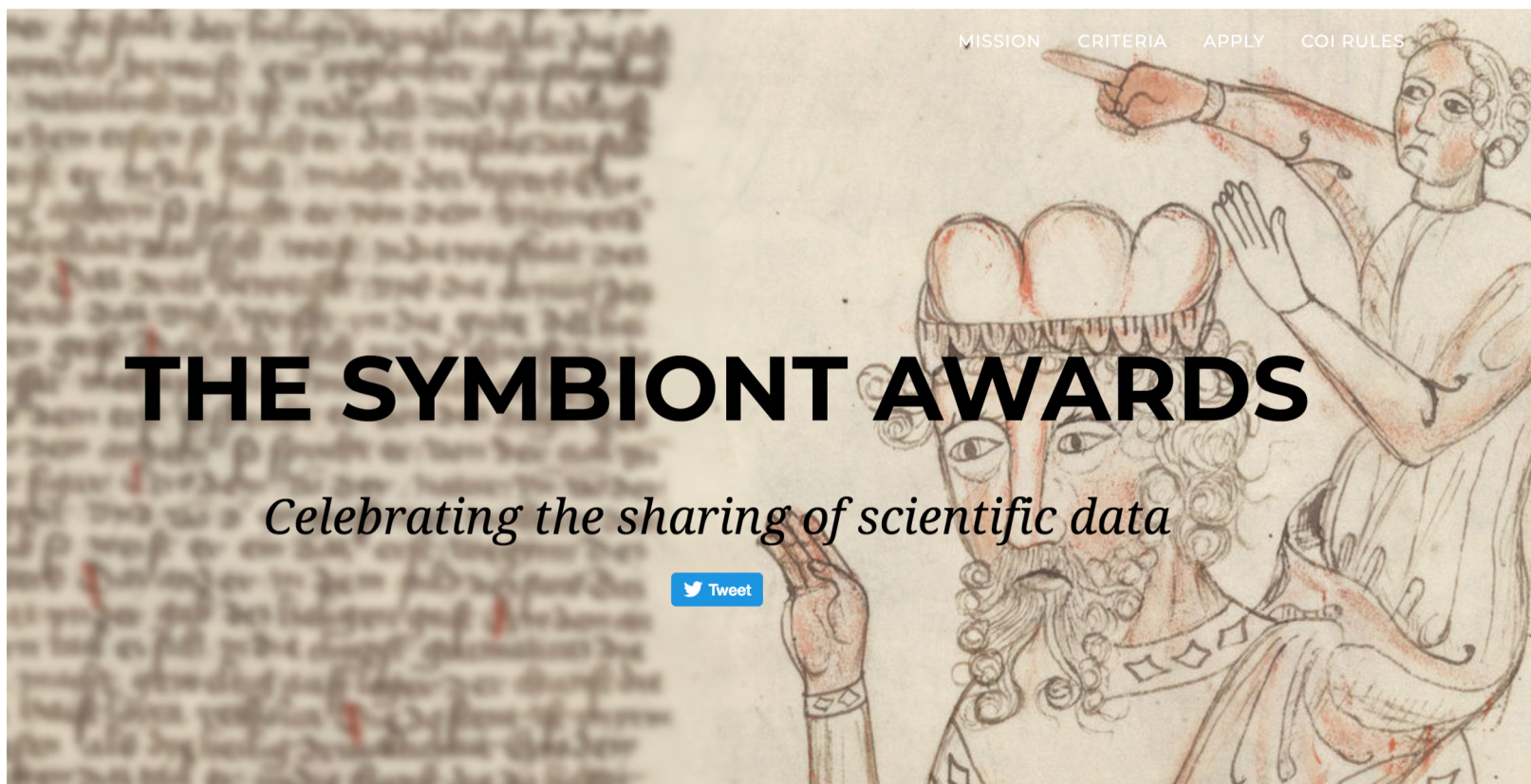
Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.



Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.



Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.

PERSPECTIVE

Assessing scientists for hiring, promotion, and tenure

David Moher^{1,2*}, Florian Naudet^{2,3}, Ioana A. Cristea^{2,4}, Frank Miedema⁵, John P. A. Ioannidis^{2,6,7,8,9}, Steven N. Goodman^{2,6,7}

Assessment of researchers is necessary for decisions of hiring, promotion, and tenure. A burgeoning number of scientific leaders believe the current system of faculty incentives and rewards is misaligned with the needs of society and disconnected from the evidence about the causes of the reproducibility crisis and suboptimal quality of the scientific publication record. To address this issue, particularly for the clinical and life sciences, we convened a 22-member expert panel workshop in Washington, DC, in January 2017. Twenty-two academic leaders, funders, and scientists participated in the meeting. As background for the meeting, we completed a selective literature review of 22 key documents critiquing the current incentive system. From each document, we extracted how the authors perceived the problems of assessing science and scientists, the unintended consequences of maintaining the status quo for assessing scientists, and details of their proposed solutions. The resulting table was used as a seed for participant discussion. This resulted in six principles for assessing scientists and associated research and policy implications. We hope the content of this paper will serve as a basis for establishing best practices and redesigning the current approaches to assessing scientists by the many players involved in that process.



We need incentives.

This study was made possible through sharing of anonymized individual participant data from the authors of all studies. We thank the authors who were contacted for this study: C Bullen and the National Institute for Health Innovation, S Gilbody, C Hewitt, L Littlewood, C van der Meulen, H van der Aa, S Cohen, M Bicket, T Harris, **the STOP GAP study investigators including Kim Thomas, Alan Montgomery, and Nicola Greenlaw**, Nottingham University Hospitals NHS Trust, NIHR programme grants for applied research, the Nottingham Clinical Trials Unit, C Polyak, K Yuhas, C Adrion, U Mansmann, G Greisen, S Hyttel-Sørensen A Barker, R Morello, K Luedtke, M Paul, D Yahav, L Chesterton, the Arthritis Research UK Primary Care Centre, and C Hanson.

Thank you.