Estimating the credibility of past research

Dr. Ian Hussey

Digitalisation of Psychology

Agenda

For today’s meeting

01 About me

02 About you

03 Purpose of the course

04 Content, organisation, assessment, etc.

About me

Postdoctoral researcher at Dept of Psy.Dig. since 2023-09

Research & teaching on meta-science

- Improving our processes of scientific knowledge production
- Scientific error detection
- Statistical analyses and where it goes wrong
- Measurement and where it goes wrong
- R: data wrangling, visualisation, and simulation studies

Thomas Herndon in 2013
Economics PhD student

He took a class where the professor
asked students to examine the results of influential papers

He chose **Reinhart & Rogoff (2010) Growth in a Time of Debt**

“Economic growth slows dramatically when the size of a country's debt rises above 90% of Gross Domestic Product”

Reinhart & Rogoff were Harvard professors + worked for the IMF

Their paper **changed the world**

EU and UK both based their economic ‘austerity’ policies on it

Ireland & Greeces were sacrificed to stop the spread of the debt crisis to the rest of Europe

4 million Irish people took on 190 Billion CHF in debt
- 48k CHF per person

Slashed our healthcare, education, infrastructure,
- These still have not recovered

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

George Osborne plunged UK into austerity due to an 'error on a spreadsheet'

An academic paper produced by two Harvard economists and relied upon by Cameron and Osborne contained serious miscalculations, it has been claimed.

by **Jack Peat** — 2022-09-22 15:30 in **Lifestyle**

Facebook Twitter Linkedin Email Whatsapp

George Osborne plunged the UK into austerity "all for nothing" due to an error on an Excel spreadsheet, according to a podcaster.

Hosts James Harkin, Andrew Hunter Murray, Anna Ptaszynski, and Dan Schreiber – from the No Such Thing As A Fish podcast – claim the UK reacted to incorrect data in an academic paper on economic growth and inflation.

Subscribe to our Newsletter

Email address:

Your e-mail address

SIGN UP

View our Privacy Policy and Terms & Conditions

More from TLE

- Watchdog calls for ticket office closures to be delayed over machine upgrades.
- Richard Curtis uses Al Pacino's Any Given Sunday speech to inspire global action on climate change
- James Cleverly holidayed with well-connected Tory lobbyist in Gibraltar
- Elevenses: Has the Left Abandoned the Working Class?
- UK recession fears grow as economy contracts
- Labour calls on Sunak to block Liz Truss's soon-to-be-published honours list
- Mick Lynch suggests RMT union will not comply with new anti-strike legislation
- Russell Brand receives standing ovation at Wembley show following sexual assault

Agenda

For today’s meeting

01 About me

02 About you - *Mentimeter*

03 Purpose of the course

04 Content, organisation, assessment, etc.

Können wir uns duzen?

Call me Ian

Agenda

For today’s meeting

01 About me

02 About you

03 Purpose of the course

04 Content, organisation, assessment, etc.

"It is simply no longer possible to believe much of the clinical research that is published,
or to rely on the judgment of trusted physicians or authoritative medical guidelines."

Dr Marcia Angell

Former editor-in-chief of

The New England Journal of Medicine

"Much of the scientific literature, perhaps half, may simply be untrue"

Dr. Richard Horton (2015)

Editor-in-chief of

The Lancet

Science & technology | Scientific malpractice

There is a worrying amount of fraud in medical research

And a worrying unwillingness to do anything about it

Feb 22nd 2023

Share

IN 2011 BEN MOL, a professor of obstetrics and gynaecology at Monash University, in Melbourne, came across a retraction notice for a study on uterine fibroids and infertility published by a researcher in Egypt. The journal which had published it was retracting it because it contained identical numbers to those in an earlier Spanish study—except that that one had been on uterine polyps. The author, it turned out, had simply copied parts of the polyp paper and changed the disease.

Listen to this story. Enjoy more audio and podcasts on iOS or Android.

0:00 / 0:00

"From that moment I was alert," says Dr Mol. And his alertness was not merely as a reader of published papers. He was also, at the time, an editor of the *European*

A few outsiders and weirdos saw the problems
by doing something the others never thought to do:

They looked.

They looked.

NULLIUS
IN
VERBA

Agenda

For today’s meeting

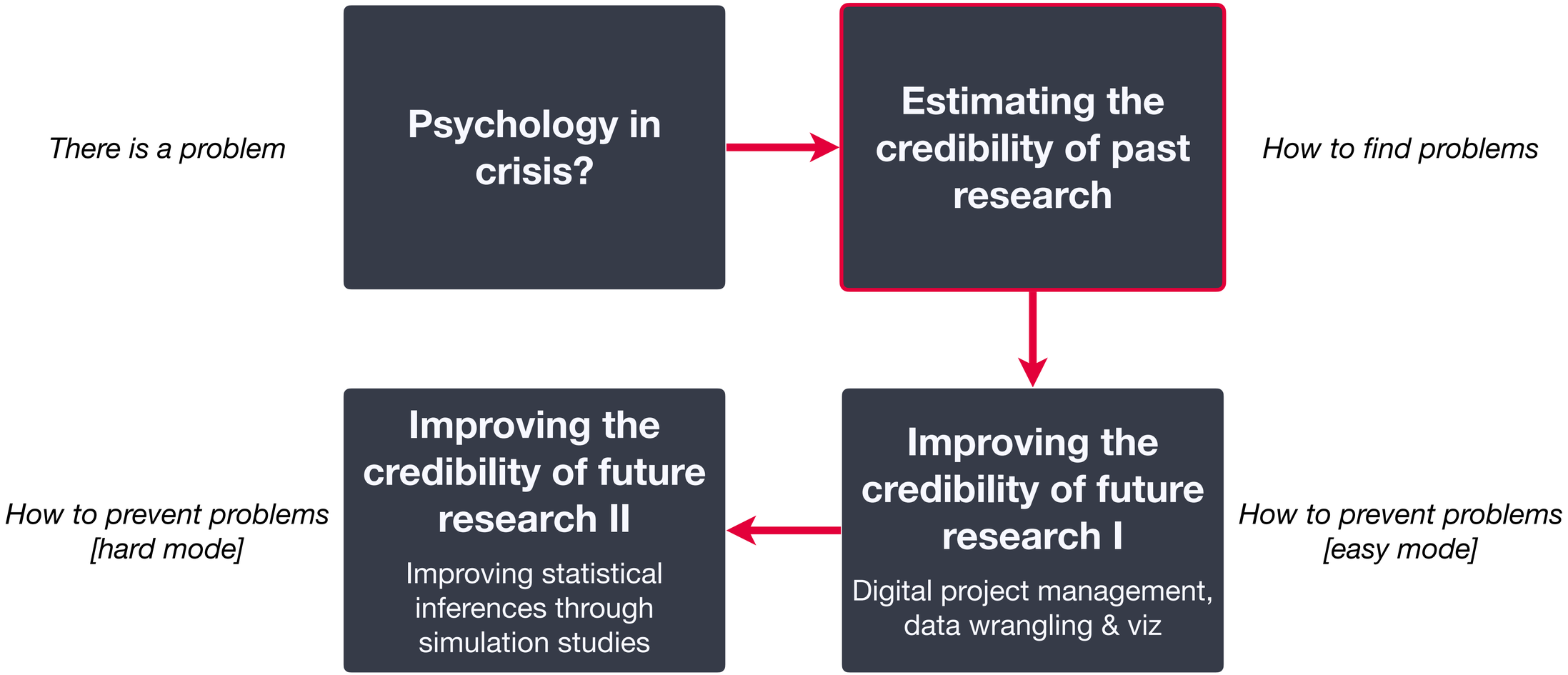
01 About me

02 About you

03 Purpose of the course

04 Content, organisation, assessment, etc.

Master's seminars I teach



Organisation & asessment

This course asks more from you than a typical seminar

Hopefully it you will also learn proportionally more

Organisation & assessment

Assessment

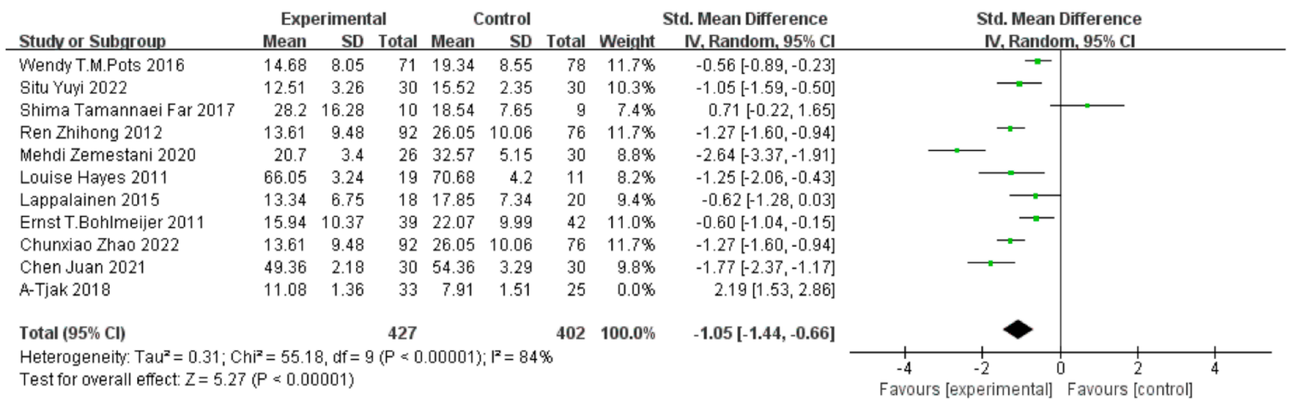
Weekly assignments on Ilias
to be completed before class

Attendance

Weekly meetings in Seminarraum 005
Bring your laptop!
You can miss max 2 sessions
Plan ahead for weekly assignments

Communication

Slack wherever possible
Discussion & problem solving encouraged!
Email if necessary: ian.hussey@unibe.ch



Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Wendy T.M.Pots 2016	14.68	8.05	71	19.34	8.55	78	11.7%	-0.56 [-0.89, -0.23]
Situ Yuyi 2022	12.51	3.26	30	15.52	2.35	30	10.3%	-1.05 [-1.59, -0.50]
Shima Tamannaei Far 2017	28.2	16.28	10	18.54	7.65	9	7.4%	0.71 [-0.22, 1.65]
Ren Zhihong 2012	13.61	9.48	92	26.05	10.06	76	11.7%	-1.27 [-1.60, -0.94]
Mehdi Zemestani 2020	20.7	3.4	26	32.57	5.15	30	8.8%	-2.64 [-3.37, -1.91]
Louise Hayes 2011	66.05	3.24	19	70.68	4.2	11	8.2%	-1.25 [-2.06, -0.43]
Lappalainen 2015	13.34	6.75	18	17.85	7.34	20	9.4%	-0.62 [-1.28, 0.03]
Ernst T.Bohlmeijer 2011	15.94	10.37	39	22.07	9.99	42	11.0%	-0.60 [-1.04, -0.15]
Chunxiao Zhao 2022	13.61	9.48	92	26.05	10.06	76	11.7%	-1.27 [-1.60, -0.94]
Chen Juan 2021	49.36	2.18	30	54.36	3.29	30	9.8%	-1.77 [-2.37, -1.17]
A-Tjak 2018	11.08	1.36	33	7.91	1.51	25	0.0%	2.19 [1.53, 2.86]
Total (95% CI)			**427**			**402**	**100.0%**	**-1.05 [-1.44, -0.66]**

Heterogeneity: Tau² = 0.31; Chi² = 55.18, df = 9 (P < 0.00001); I² = 84%
Test for overall effect: Z = 5.27 (P < 0.00001)

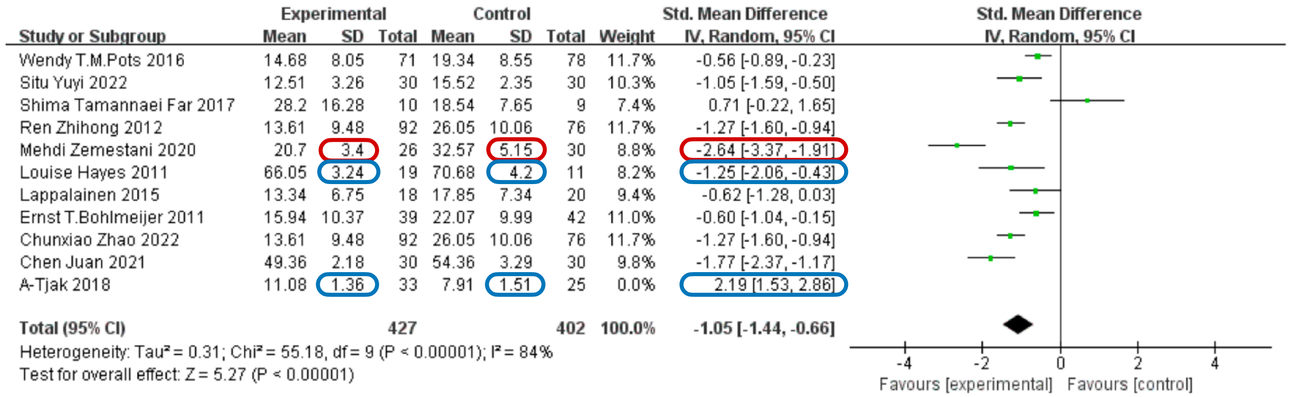
Fig. 4 Effect of ACT on depression in patients with depressive disorders after sensitivity analysis

This class will teach you skills and intuitions to estimate **trustworthiness** & **credibility**

- eg Standardized Effect sizes that are implausibly large
- SDs that are implausibly small

To demonstrate this in my 2023 class, I checked one (1) article

Zhao et al. (2023) Effect of Acceptance and Commitment Therapy for depressive disorders: a meta-analysis



Study or Subgroup	Experimental Mean	Experimental SD	Experimental Total	Control Mean	Control SD	Control Total	Weight	Std. Mean Difference IV, Random, 95% CI
Wendy T.M.Pots 2016	14.68	8.05	71	19.34	8.55	78	11.7%	-0.56 [-0.89, -0.23]
Situ Yuyi 2022	12.51	3.26	30	15.52	2.35	30	10.3%	-1.05 [-1.59, -0.50]
Shima Tamannaei Far 2017	28.2	16.28	10	18.54	7.65	9	7.4%	0.71 [-0.22, 1.65]
Ren Zhihong 2012	13.61	9.48	92	26.05	10.06	76	11.7%	-1.27 [-1.60, -0.94]
Mehdi Zemestani 2020	20.7	3.4	26	32.57	5.15	30	8.8%	-2.64 [-3.37, -1.91]
Louise Hayes 2011	66.05	3.24	19	70.68	4.2	11	8.2%	-1.25 [-2.06, -0.43]
Lappalainen 2015	13.34	6.75	18	17.85	7.34	20	9.4%	-0.62 [-1.28, 0.03]
Ernst T.Bohlmeijer 2011	15.94	10.37	39	22.07	9.99	42	11.0%	-0.60 [-1.04, -0.15]
Chunxiao Zhao 2022	13.61	9.48	92	26.05	10.06	76	11.7%	-1.27 [-1.60, -0.94]
Chen Juan 2021	49.36	2.18	30	54.36	3.29	30	9.8%	-1.77 [-2.37, -1.17]
A-Tjak 2018	11.08	1.36	33	7.91	1.51	25	0.0%	2.19 [1.53, 2.86]
Total (95% CI)			**427**			**402**	**100.0%**	**-1.05 [-1.44, -0.66]**

Heterogeneity: Tau² = 0.31; Chi² = 55.18, df = 9 (P < 0.00001); I² = 84%
Test for overall effect: Z = 5.27 (P < 0.00001)

Fig. 4 Effect of ACT on depression in patients with depressive disorders after sensitivity analysis

This class will teach you skills and intuitions to estimate **trustworthiness** & **credibility**

- eg Standardized Effect sizes that are implausibly large
- SDs that are implausibly small

To demonstrate this in my 2023 class, I checked one (1) article

Zhao et al. (2023) Effect of Acceptance and Commitment Therapy for depressive disorders: a meta-analysis

In 2024, I again checked one (1) article

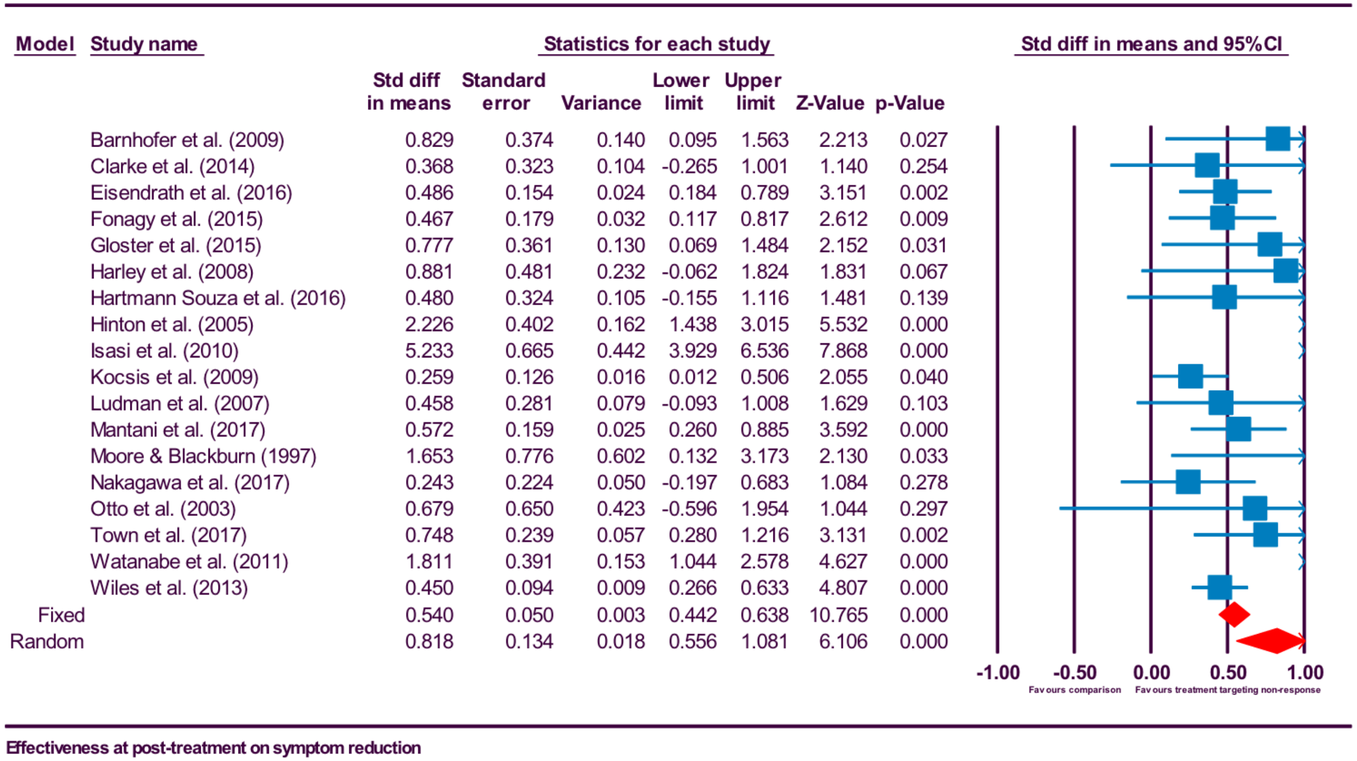
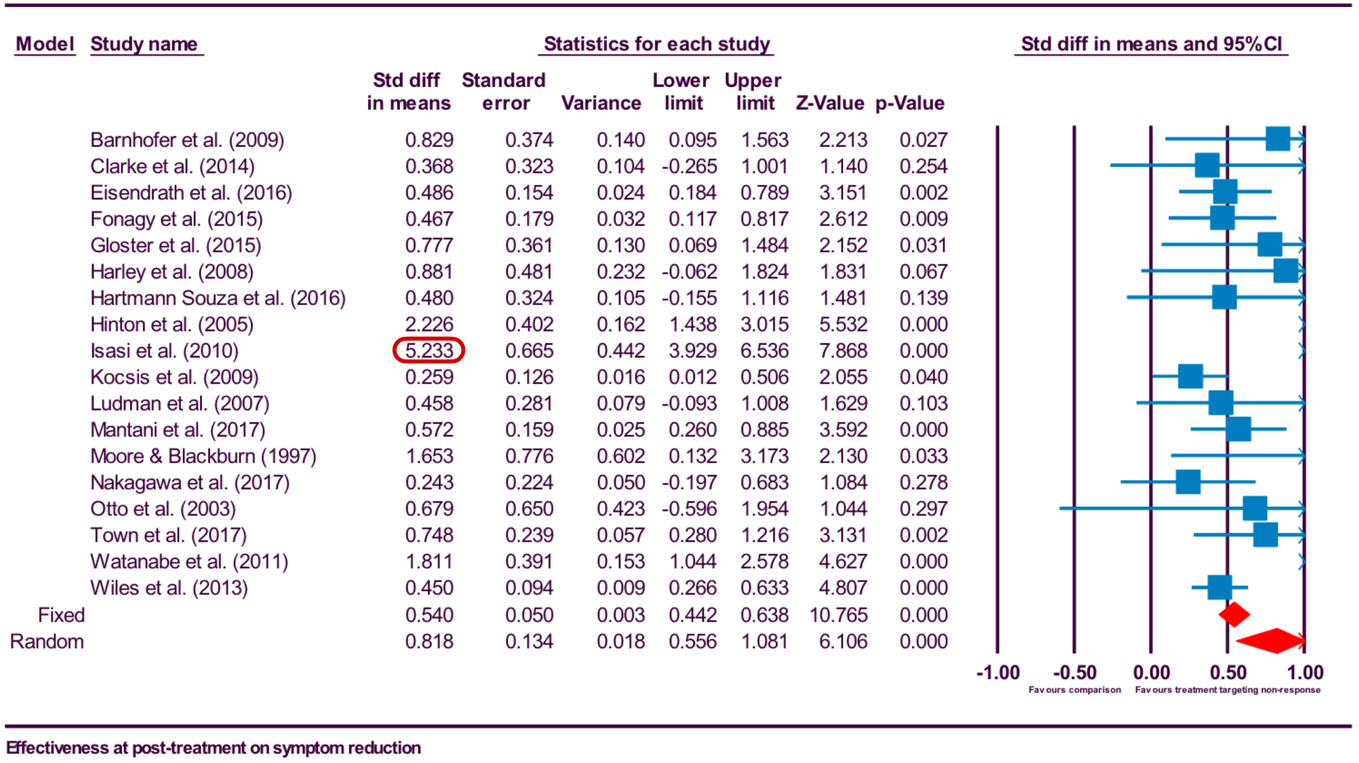


Fig. 2. Forest plot of the meta-analysis on primary outcomes (symptoms).

Gloster et al. (2020) Treating treatment non-responders: A meta-analysis of randomized controlled psychotherapy trials



Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Barnhofer et al. (2009)	0.829	0.374	0.140	0.095	1.563	2.213	0.027
	Clarke et al. (2014)	0.368	0.323	0.104	-0.265	1.001	1.140	0.254
	Eisendrath et al. (2016)	0.486	0.154	0.024	0.184	0.789	3.151	0.002
	Fonagy et al. (2015)	0.467	0.179	0.032	0.117	0.817	2.612	0.009
	Gloster et al. (2015)	0.777	0.361	0.130	0.069	1.484	2.152	0.031
	Harley et al. (2008)	0.881	0.481	0.232	-0.062	1.824	1.831	0.067
	Hartmann Souza et al. (2016)	0.480	0.324	0.105	-0.155	1.116	1.481	0.139
	Hinton et al. (2005)	2.226	0.402	0.162	1.438	3.015	5.532	0.000
	Isasi et al. (2010)	5.233	0.665	0.442	3.929	6.536	7.868	0.000
	Kocsis et al. (2009)	0.259	0.126	0.016	0.012	0.506	2.055	0.040
	Ludman et al. (2007)	0.458	0.281	0.079	-0.093	1.008	1.629	0.103
	Mantani et al. (2017)	0.572	0.159	0.025	0.260	0.885	3.592	0.000
	Moore & Blackburn (1997)	1.653	0.776	0.602	0.132	3.173	2.130	0.033
	Nakagawa et al. (2017)	0.243	0.224	0.050	-0.197	0.683	1.084	0.278
	Otto et al. (2003)	0.679	0.650	0.423	-0.596	1.954	1.044	0.297
	Town et al. (2017)	0.748	0.239	0.057	0.280	1.216	3.131	0.002
	Watanabe et al. (2011)	1.811	0.391	0.153	1.044	2.578	4.627	0.000
	Wiles et al. (2013)	0.450	0.094	0.009	0.266	0.633	4.807	0.000
Fixed		0.540	0.050	0.003	0.442	0.638	10.765	0.000
Random		0.818	0.134	0.018	0.556	1.081	6.106	0.000

Fig. 2. Forest plot of the meta-analysis on primary outcomes (symptoms).

In 2024, I again checked one (1) article

& I found serious errors

surement tools. The standardized mean difference is calculated by the following equation:

SMD
= Difference in mean outcome between groups
/Change of SD of outcome between groups.

- Their definition of SMD is wrong
- Effect size cannot be recreated from the numbers in the original article

Gloster et al. (2020) Treating treatment non-responders: A meta-analysis of randomized controlled psychotherapy trials

Readings

- Bernal (1939) [read]
- Van Noorden (2023) How many clinical trials can't be trusted? [skim]
- Simply Science newsletter (2023) There is a worrying amount of fraud in medical research [skim]
- Smith (2021) Time to assume that health research is fraudulent until proven otherwise? [skim]

Assignment (2000-5000 characters)

- Sometimes its easy to lose perspective on the big questions, such as "why do scientists do science?" We are often told that we do science to discover truth. If this is the only reason errors and fraud would be rare, but there is growing concern they are not (Van Noreen, 2023; Simply Science, 2023; Smith, 2021). Bernal (1939) describes other reasons why scientists do science. Which of these reasons have been emphasised in your own studies so far, and which have not? Do you agree with Bernal?

Questions?