



Increase value and reduce waste in research on psychological therapies

Ioana A. Cristea^{a,b,*}, Florian Naudet^c

^a Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Republicii 37 Street, Cluj-Napoca, Romania

^b Meta-Research Innovation Center at Stanford (METRICS), Stanford University, Stanford, CA 94305, California, USA

^c University Rennes, CHU Rennes, Inserm, CIC 1414 (Centre d'Investigation Clinique de Rennes), F- 35000, Rennes, France



ARTICLE INFO

Keywords:

Research waste
Open science
Reproducibility
Transparency
Psychological treatment
Psychotherapy
meta-Research
Risk of bias
Publication bias
Data sharing

ABSTRACT

A seminal Lancet series focused on increasing value and reducing waste in biomedical research, providing a transferrable template to diagnose problems in research. Our goal was to document how some of these sources of waste apply to mental health and particularly psychological treatments research. We synthesize and critically evaluate empirical findings in relation to four major sources: i) defining research priorities; ii) research design, methods and analysis; iii) accessibility of research information; iv) accuracy and usability of research reports. We demonstrate that each source of waste considered is well-represented and amply documented within this field. We describe hype and insufficient consideration of what is known in defining research priorities, persistent risk of bias, particularly due to selective outcome reporting, for psychotherapy trials across mental disorders, intellectual and financial biases, direct and indirect evidence of publication bias, largely inexistent adoption of data sharing, issues of multiplicity and fragmentation of data and findings, and insufficient adoption of reporting guidelines.

We expand on a few general solutions, including supporting meta-research, properly testing interventions to increase research quality, placing open science at the center of psychological treatment research and remaining vigilant particularly regarding the strains of research currently prioritized, such as experimental psychopathology.

In 2014 a seminal Lancet series focused on increasing value and reducing waste in biomedical research. In the prefacing commentary, M. R. Macleod et al. (2014) start from a case example both straightforward and striking: out of 1575 reports about cancer prognostic markers published in 2005, 96% claimed to have discovered at least one significant prognostic marker (Kyzas, Denaxa-Kyza, & Ioannidis, 2007). Yet only a minority of markers were replicated in subsequent research, with adoption in clinical practice similarly limited (Anderson, 2010). M. R. Macleod et al. (2014) maintain that such a pattern of initially promising findings that nonetheless fail to engender tangible healthcare improvements is the rule rather than the exception across biomedical research. Avoidable waste in the production and reporting of biomedical research had previously been estimated at a staggering 85% (Chalmers & Glasziou, 2009). The papers in the Lancet series explore possible reasons for the failure to produce transformative and useful research, with each examining a potential source of “research waste”. The reports discuss whether i) decisions about what research to fund are based on questions relevant to users of research (Chalmers et al., 2014); ii) appropriate research designs, methods and analyses are being employed (Ioannidis et al., 2014); iii) research regulation and

management is efficient (Salman et al., 2014); iv) research information is fully accessible (Chan et al., 2014) and finally v) research reports are unbiased and useable (Glasziou et al., 2014). Each report delineates various steps and initiatives for reducing waste and increasing value, with several of these subsequently expanded in similar contributions, such as the “Manifesto for reproducible science” (Munafò et al., 2017).

Beyond the content of the included papers, the Lancet series sparked a movement for reducing waste and increasing value in research. This is reflected in initiatives like the REWARD Alliance (<http://rewardalliance.net/>), a dedicated platform gathering documentation, information, resources and a string of initiatives, such as working groups, prizes, blogs and other tools. More generally, the Lancet series papers coalesce in an exportable framework for diagnosing problems with research conducted in other healthcare domains. Though examples in the reports occasionally touch on mental health or psychotherapy research, their main emphasis rests with biomedical research and treatments. Still, a closer examination of specific subfields allows detecting particular problems and suggesting tailored solutions. Therefore, in this opinion piece, our goal is to document how some of the sources of waste identified apply to research in mental health, and

* Corresponding author. Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Republicii 37 Street, 400015, Cluj-Napoca, Romania.
E-mail addresses: ioana.alina.cristea@gmail.com (I.A. Cristea), floriannaudet@gmail.com (F. Naudet).

in particular on psychological treatments.

1. Are research decisions based on questions relevant to users of research?

Though any research aimed at finding, testing or understanding the nature and treatment of mental disorders is potentially valuable, limited resources set constraints and impose defining priorities. Mental health funding is usually limited (Hazo et al., 2016), though with significant discrepancies across countries (Hazo et al., 2017). Funding calls and reports make it clear certain topics are prioritized. For instance, the MQ report on mental research funding in the UK showed that about a quarter (23.7%) of the annual mental health funding goes into “basic psychology and neuroscience” and about half goes into research on underpinning (i.e., brain structure and functioning in healthy individuals) and aetiology (MQ, 2019). The Wellcome Trust, one of the largest non-profit funders, recently announced a commitment of £200 million for research on discovering and mobilizing treatment mechanisms (Thielking, 2019).

Similar research priorities were described in a Lancet Psychiatry Commission (Holmes et al., 2018), which delineated an agenda for research on psychological treatments. The fundamental premise of this programmatic document was that psychological treatment research had strayed away from the mechanistically informed approaches that were the foundation for some of the current effective treatments, (e.g., exposure and response prevention for obsessive-compulsive disorder), to instead reorient itself towards modifying or adapting existing treatments. Owing to this shift, the authors contend, even if effective psychological treatments are available across a variety of settings and populations, there is little knowledge about their mechanisms of action. In response, the authors of the Commission champion the burgeoning field of experimental psychopathology, broadly encompassing studies in which the experimental manipulation of a purported mechanism of treatment leads to symptom change. This field promises to identify “key processes that maintain or change aspects of psychopathology” (Holmes et al., 2018) (p. 244), as well as to determine which processes are modifiable and could consequently represent appropriate treatment targets.

Similarly to research on prognostic markers for various cancers, experimental psychopathology unlocks a multitude of exciting possibilities. Still, analogously to prognostic markers, many modest findings coming from experimental psychopathology, with as yet minimal treatment implications, have been replaced by hype. Several of the most touted and oft-cited developments, mentioned in the Commission as prototypical research in this field, have not been replicated in subsequent, independently conducted, studies or had little impact for translation to clinical practice. One revealing example is cognitive bias modification, a class of interventions developed in the laboratory and broadly encompassing techniques for the direct manipulation of a target cognitive bias with the goal of obtaining symptom change. Attention bias modification (ABM) interventions, the most prominent example in this class, are premised on experimental studies supporting the existence of a causal relationship between attentional bias and anxiety (C. MacLeod & Mathews, 2012; Van Bockstaele et al., 2014). Based on these largely laboratory-based studies and early small trials, ABM interventions were hailed as “having passed the proof-of-concept stage” (C. MacLeod, 2012) (p.118), effective in reducing dysfunctional anxiety (C. MacLeod & Mathews, 2012) and nothing short of a “cognitive vaccine” for depression (Browning, Holmes, Charles, Cowen, & Harmer, 2012). Yet in parallel with growing excitement, larger clinical trials accrued and mostly produced negative results. Subsequent meta-analyses showed that ABM and other bias modification approaches had limited, if any, effects, for anxiety and depression outcomes, in both adult (Cristea, Kok, & Cuijpers, 2015) and young populations (Cristea, Mogoase, David, & Cuijpers, 2015). Moreover, effects were particularly small in clinical populations and when the intervention was delivered

in non-laboratory settings, the very contexts most relevant for translation to practice. Ironically, not only did translation from the experimental studies to treatments useful to patients prove largely a failure, but even the notion of surpassing the proof-of-concept stage has recently been questioned. One meta-analysis failed to find evidence that clinically anxious individuals enrolled in ABM trials displayed attentional bias to begin with (Kruijt, Parsons, & Fox, 2019).

Another illustrative example is memory reconsolidation, the notion that postretrieval learning can be employed to modify or even erase existing memory traces, effectively affording “rewriting” of existing memories. This theory stems from basic research claiming to demonstrate reconsolidation effects, first in animals (Nader, Schafe, & Le Doux, 2000), and subsequently in humans (Schiller et al., 2009; Walker, Brakefield, Hobson, & Stickgold, 2003). Should the initial laboratory findings prove reproducible and translatable, treatment implications would be immense, particularly in disorders where intrusive memories play a central part, like post-traumatic stress disorder (PTSD).

Yet both reproducibility and translation are doubtful. The initial experimental results demonstrating memory reconsolidation effects in humans failed both procedural and conceptual replication by independent groups (Hardwicke, Taqi, & Shanks, 2016; Kindt & Soeter, 2013). In terms of translation, a recent randomized trial attempted to mitigate future post-traumatic symptoms in motor vehicle collision survivors in an emergency setting (Iyadurai et al., 2018). The intervention, founded on a theoretical mix between memory consolidation and reconsolidation, consisted of a brief trauma reminder followed by playing the computer game Tetris. It had already been tested in the laboratory on healthy participants who were induced symptoms labeled as trauma-analogue, by means of watching a traumatic film, and it had showed promising effects in reducing memory flashbacks over the course of a week (Holmes, James, Kilford, & Deerprouse, 2010). The subsequent trial on trauma victims showed a similar reduction in flashbacks over 1 week, on the basis of which the authors declared the intervention a “compelling translation of previous laboratory findings” (p.680) and likened it to a “cognitive therapeutic vaccine”. In a commentary, we disputed this characterization, questioning the magnitude and stability of the observed effects (Cristea, Naudet, Shanks, & Hardwicke, 2018), which were largely influenced by outliers, reduced in a more robust analysis, and, critically, not observed at 1-month follow-up. Furthermore, the stakes of a successful translation would involve not just reducing post-trauma flashbacks, but delaying or avoiding the onset of PTSD altogether. To this purpose, a larger, phase III trial would be necessary, employing a heterogeneous sample of trauma victims, pre-screened for having a high level of post-traumatic flashbacks, with reliably measured primary outcomes and longer follow-up durations, including for monitoring adverse effects. It is currently unclear whether conducting such a costly and logistically complex trial is a judicious investment. Its foundation would consist of non-replicated experimental findings, favorable results observed in healthy participants in a laboratory context that has little in common with real trauma, and transient effects of disputable magnitude in one proof-of-concept trial on victims of one type of trauma. Given its considerable risk of producing null results and not meaningfully informing clinical practice, such a trial would rather amount to research waste. Similar doubts apply to other promising experimental psychopathology findings.

When confronted with replication or translation failures, researchers often hypothesize potential moderators or boundary conditions to account for small, elusive or inconsistently observed effects. This is certainly a distinct possibility in some cases, but in others the original results were probably artefacts. Still, moderators or boundary conditions are usually inferred post-hoc and would also need to be examined empirically and prospectively (Cristea, 2018a; Hardwicke & Shanks, 2016). More importantly, it remains questionable whether priorities in research on psychological interventions should coalesce around effects of this size and stability. Of course, it is impossible to

know *a priori* whether a research idea will lead to massive progress or to fleeting, non-replicable, or difficult to translate effects. As is often stated, absence of evidence is not evidence of absence. We are not denying the importance of attempting to identify treatment mechanisms, nor are we arguing that exploring novel ideas should not be encouraged or is in itself a source of waste. However, we are concerned waste might result from two other sources. First, the unjustified hype that usually accompanies research in areas like experimental psychopathology might be counterproductive. As with the drug development pipeline, many of the exciting pre-clinical or early clinical findings prove to be dead-ends (Contopoulos-Ioannidis, Ntzani, & Ioannidis, 2003). If they are conveyed with forceful and often grandiose language, like by analogy with vaccines, high and unwarranted expectations are set up in other interested researchers and the general public. The untoward consequences might include reticence in researchers, reviewers and journal editors to credit and publish negative findings, as well as premature implementation of an insufficiently tested intervention, which may be ineffective or even harmful. Second, continued efforts to “rescue” a hypothesis in the face of accumulating negative results, simply because initial findings were promising, is also counterproductive. Instead, resource limitations would dictate reorienting efforts to other ideas.

Another source of waste in setting research priorities is ignoring what is already known or being researched (Chalmers et al., 2014). Often, research questions are grounded in a narrative synthesis, which facilitates cherry-picking ideas to support preferred hypotheses. One possible preventing strategy is requiring that a systematic review and, if feasible, a meta-analysis preclude any investment or funding decisions, so as to ensure that the question has not already been answered satisfactorily. Determining which questions have value is more challenging, since allocating resources to any priority comes at the cost of divesting them from alternatives. For example, the focus on mechanistic approaches to psychological treatments could distract from other, pervasive, determinants of mental disorders, like poverty (Deighton et al., 2019), income inequality (Patel et al., 2018) or living in a conflict zone (Jayasuriya, Jayasuriya, Tay, & Silove, 2016). Specifically, proposing a simple behavioral intervention like the Tetris game to tackle intrusive memories in refugees from warzones like Syria (Holmes et al., 2017) might deflect from the disabling and enduring consequences of war-related trauma (Silove et al., 2014).

Waste in setting research priorities also ensues when users of research are not involved (Chalmers et al., 2014). Patient input is crucial regarding key outcomes, type of treatment preferred (i.e., psychological, pharmacological), or the design (i.e., superiority, equivalence, non-inferiority) that answers the most important question from a patient perspective (Holmes et al., 2018). Pragmatically, there are direct and immediate benefits of patient involvement in trials. For example, a meta-analysis showed that patient and public involvement, especially for people with lived experience, improved participant enrolment (Crocker et al., 2018). Patient preferences have still not been sufficiently probed in psychological treatment research. Most analyses so far have focused on defining target outcomes that matter to patients (Cuijpers, 2019). But whether patients favour many of the currently prioritized mental health topics remains an open question. For instance, how do deeply traumatized refugees feel about being offered a simple behavioural intervention like the Tetris game? What we are advocating for is a vision of research as an enterprise that advances collaboratively with patients instead of the traditional paternalistic view of research done *for* patients.

2. Appropriate research design, methods, and analysis?

A few years ago, we provocatively proposed a set of guidelines on how to prove a psychological treatment is effective even when in reality it is not (Cuijpers & Cristea, 2016). We had intended these as a tongue-in-cheek recipe of how not to design, conduct and report trials of

psychological interventions. We described three categories of behaviours, ranging from those clearly inadequate to those problematic under certain circumstances. The first one involved exploiting the design, implementation and reporting of a trial so as to “prettify” the results and make an intervention appear more effective. Examples included: i) not concealing the allocation of the participants from the research staff who could, unwittingly or not, use this information to influence the delivery of the intervention; ii) not using blind assessors of outcome, hence allowing them to be influenced by knowledge of assignment and possibly judge participants in the intervention group as more improved; iii) including a stack of outcome measures and selectively reporting the outcomes with positive findings; iv) reporting analysis on completers instead of all randomized participants (i.e., intent-to-treat), since completers are most likely to have benefited from treatment and therefore remained in the study; and v) holding back the publication of negative results. The second category encompassed choices in research design that can be appropriate in the early stages of research, but become questionable when studies with these characteristics dominate the evidence base, such as: i) using waiting list controls; ii) employing small samples; iii) not comparing the intervention with others known to be effective. Finally, the third category involved behaviours in themselves not inadequate, and which many researchers view as the mark of competence and commitment in developing and administering interventions. These include various ways of enhancing users' expectations, like expressing your own confidence in the intervention, using a variety of methods to convey trust and enthusiasm, such as writing targeted books, giving lectures at conferences and events with professionals, giving interviews in the media, or proposing anecdotal evidence like patient testimonials and success stories. Though not intrinsically inappropriate, the intellectual investment in an approach could function as a bias, reducing vigilance in detecting inappropriate choices in design, analysis and reporting, particularly in the presence of findings favorable for the intervention. Along the same lines, Meichenbaum and Liliencfeld (2018) proposed a 19-item “psychotherapy hype checklist”, describing warning signs that an intervention's efficacy has been significantly exaggerated, grouped under promotion /marketing, and research evidence.

Are these counter-suggestions illustrative of the reality “on the ground” in research on psychological interventions? Meta-epidemiological investigations that systematically look at patterns in collections of studies, like systematic reviews, surveys, meta-analyses or umbrella reviews (i.e., systematic reviews of systematic reviews) appear to support this thesis, indicating endemic problems. For clinical trials, potential markers of questionable practices in the design, methods or analysis are represented by the risk of bias (RoB) domains (Higgins et al., 2011). These include: (i) adequate sequence generation; (ii) concealment of allocation; (iii) blinding of outcome assessors; (iv) dealing with incomplete data; and (v) selective reporting. Assessing trial risk of bias is standard practice in a meta-analysis, offering us a snapshot of the state of the field. Across meta-analyses, a consistent finding is that there are few trials of psychological interventions with low RoB across all domains. High or uncertain RoB seems to be ubiquitous across the whole range of treatment approaches and disorders, such as cognitive behaviour therapy (CBT) for major depression and anxiety disorders (Cuijpers, Cristea, Karyotaki, Reijnders, & Huibers, 2016), interpersonal therapy for mental disorders (Cuijpers, Donker, Weissman, Ravitz, & Cristea, 2016), virtual reality-based therapies (Fodor et al., 2018), or psychological treatments in general for depression (Cuijpers, Karyotaki, Reijnders, & Ebert, 2019), social anxiety (Mayo-Wilson et al., 2014), panic disorder (Pompili et al., 2016), obsessive-compulsive disorder (Skapinakis et al., 2016), PTSD (Bisson, Roberts, Andrew, Cooper, & Lewis, 2013), schizophrenia (Bighelli et al., 2018), bipolar disorder (Chatterton et al., 2017), borderline personality disorder (Cristea et al., 2017) or eating disorders (Slade et al., 2018). An inverse relationship between study quality, conceptualized as low overall RoB, and the magnitude of effects was consistently shown either

as a direct relationship in meta-regression analyses (Barth et al., 2013; Cristea et al., 2017), or as a considerable effect size reduction in analyses restricted to studies with low RoB (Cuijpers, Cristea, et al., 2016; Cuijpers et al., 2019).

A particularly critical and difficult to assess bias is selective outcome reporting, which cannot be detected in the absence of complete and accurate information about what investigators planned to do and why. Prospective registration of trial protocols in publicly accessible registries is therefore an indispensable prerequisite. Selective reporting bias is closely linked with another source of waste, namely the usability of research reports and protocols, discussed in a further section. It encompasses an array of inappropriate reporting behaviours, ranging from not reporting certain planned outcomes at all, to selecting which of the measurements of an outcome to report, to modifying the prioritization of planned outcomes or measurements (i.e., downgrading a primary outcome to secondary or upgrading a secondary outcome to primary). Crucially, reporting decisions are usually based on the direction, magnitude or significance of an effect estimate, so as to elevate positive results, and minimize or eclipse negative ones. As a consequence, the intervention appears more effective or less harmful than it really is. Selectively reported trials are included in research synthesis and have a lasting distorting effect on the evidence base. This pernicious bias seems to be rampant in research on psychological interventions. A study of selective reporting in depression trials found that only 20% of the CBT trials had been properly registered, and only around 14% both properly registered and reported (Shinohara et al., 2015). Another investigation of psychotherapy trials published in five top-tier dedicated journals in clinical psychology showed that under a quarter of the surveyed trials had been prospectively registered and that this proportion dropped to just under 12% when the accuracy of registration (i.e., no ambiguities regarding the primary outcome measure or the timeframe of assessment) was taken into account (Bradley, Rucklidge, & Mulder, 2017). Under 5% of trials were judged free of selective outcome reporting, i.e., had no discrepancies between the registration and the report.

Another ubiquitous and often inappropriate research design choice is resorting to weak control groups, such as the waiting list. Based largely on comparisons to this condition, estimates of effects for psychological interventions can appear very large. For instance, a meta-analysis of CBT (Cuijpers, Cristea, et al., 2016) found very large effects when compared to waiting list controls, ranging from 0.85 for generalized anxiety disorder to 0.96 for panic disorder, and 0.98 for social anxiety disorder and, major depressive disorder. For depression, meta-analyses indicated that participants assigned to the waiting list showed diminished response in terms of symptom reduction, and that this condition was significantly inferior to other types of control, like treatment as usual or placebo (Barth et al., 2013; Furukawa et al., 2014; Khan, Faucett, Lichtenberg, Kirsch, & Brown, 2012). Some researchers even likened the waiting list to a placebo (Furukawa et al., 2014). For social anxiety disorder, a meta-analysis (Steinert, Stadter, Stark, & Leichsenring, 2017) showed a pooled within group effect (Hedges' g) of 0.13 for the waiting list arms, considerably smaller than that of active treatments (Hedges' g = 0.88). In another meta-analysis (Fodor et al., 2018) of virtual reality-based intervention for anxiety outcomes across multiple anxiety disorders, the effects versus waiting list controls were very large (Hedges' g = 0.90). We argued that the waiting list is an inadequate benchmark for estimating the effectiveness of psychotherapy, at least for depression (Cristea, 2018b). Due to the disadvantages this type of control carries in terms of possible harm to patients and an overestimation of an intervention's effects, we contend it should be used sparingly (Cristea, 2018b), possibly as part of designs that attempt to maximize patients' time while waiting for therapy (Lovell et al., 2017) or that allow them to refuse waiting list assignment and still remain in the trial (i.e., equipoise-stratified). At least for the more prevalent mental disorders discussed above (depression, anxiety), there is an excess of trials with waiting list arms. Supportive, non-

specific or minimal contact interventions are more appropriate control options for newly developed psychological treatments.

Small study effects, the tendency of interventions to appear more beneficial in smaller studies (Higgins & Green, 2011), are also pervasive in research on psychological interventions. An umbrella review of meta-analyses of randomized controlled trials of psychotherapy for a variety of outcomes including symptoms of mental disorders, distress and physical outcomes, found evidence of small study effects for about 30%–40% of the included meta-analyses, depending on the method used (Dragioti, Karathanos, Gerdle, & Evangelou, 2017). Conversely, an umbrella review of psychological interventions for pain reduction across a range of conditions, most of them physical, did not report evidence of small study effects (Markozannes et al., 2017).

Risk of bias, weak control groups and small study effects are not circumscribed to a few trials and probably act synergistically in distorting estimates. In an interesting analysis, Cuijpers et al. (2019) tried to estimate the “bias-free” effectiveness of psychological treatments for depression, by sequentially eliminating each source of bias and recalculating the pooled treatment effect. They started from an initial unadjusted estimate (Hedges' g) of 0.63 (325 comparisons), which dropped to 0.51 (179 comparisons) when comparisons with the waiting list were excluded, 0.38 (71 comparisons) when only studies with low risk of bias across domains were included, and respectively 0.31 (84 comparisons) when a method was used to attempt to correct the pooled effect for the presence of small study effects.

Other biases are also prevalent, though their distorting influence of outcomes is difficult to ascertain. Contamination by researcher allegiance, the belief of the investigators in the superiority of one of the treatments studied, has long been debated (Munder, Fluckiger, Gerger, Wampold, & Barth, 2012). Over 60% of the trials included in meta-analyses of psychotherapy were appraised as allegiant in a systematic evaluation (Dragioti, Dimoliatis, & Evangelou, 2015). However, defining and reliably assessing which researcher behaviors constitute allegiance remain challenging aspects (Yoder, Karyotaki, Cristea, van Duin, & Cuijpers, 2019). Still, some meta-analyses of psychological interventions, using diverse assessment procedures, reported an association between the presence of allegiant investigators and outcomes (Munder, Gerger, Trelle, & Barth, 2011; D. T. Turner, van der Gaag, Karyotaki, & Cuijpers, 2014). A survey of meta-analyses of RCTs of psychotherapy revealed that effects were larger by almost 30% for trials involving allegiant investigators (Dragioti, Dimoliatis, Fountoulakis, & Evangelou, 2015). However, allegiance effects are observational and might also reflect reverse causality: investigators become allegiant because they are exposed to treatments that are truly more effective (Wilson, Wilfley, Agras, & Bryson, 2011). Furthermore, other variables such as therapist competence or treatment integrity could confound the relation between allegiance and outcomes. For instance, trials with allegiant investigators did not result in more favorable outcomes when treatment integrity was assessed in a survey of meta-analyses (Dragioti, et al., 2015). Similarly, a randomized trial comparing an IPT and a CBT intervention for binge eating disorders found no evidence of allegiance bias though ostensibly allegiant investigators were involved (Wilson et al., 2011), a result the authors partly attribute to high treatment integrity. Finally, investigator background, such as specialization or affiliation (i.e., a biomedical or psychosocial concentration), might also engender an intellectual bias, though we did not find corroborating evidence in a meta-analysis of psychotherapy and pharmacotherapy for depression (Cristea, Gentili, Pietrini, & Cuijpers, 2017a).

Beyond the intellectual biases, more direct, financial conflicts of interests (COIs) in relationship to psychological interventions may have remained largely under the radar (Cristea & Ioannidis, 2018). We charted a variety of possibly applicable financial COIs, ranging from commercial companies developing and testing psychological interventions, a situation akin to the pharmaceutical industry, to less straightforward cases of potential financial gain, such as involvement with academic spin-offs, temporary participations in commercial ventures,

for-profit support activities like training, and involvement with foundations or non-profits. There is a dearth of systematic evaluations as to whether these strains of financial COIs distort effectiveness estimates for psychological interventions.

Another related bias, sponsorship bias, refers to the notion that industry funding of trials is associated with findings favoring the sponsor's product. The study of this bias was traditionally confined to trials of pharmacological treatments, but we reported corroborating evidence in head-to-head RCTs of psychotherapy versus pharmacotherapy (Cristea, Gentili, Pietrini, & Cuijpers, 2017b). Industry-funded depression trials tended to report more favorable results for pharmacotherapy versus psychotherapy, a pattern not present in non-industry funded trials. Trials where authors had financial COIs linked to the pharmaceutical industry showed a similar pattern (Cristea et al., 2017a; 2017b).

3. Fully accessible research information?

The lack of full and accurate information about a completed study, ranging from its planning and protocol to its full set of results, is another critical source of research waste. Two major categories of problems refer to (1) findings that are never published, mostly because they would reflect badly on an intervention, its developers, sponsors or sometimes an entire research field (i.e., publication bias) and (2) data underlying reported findings not being made accessible to other researchers or the public (i.e., data sharing).

Most evidence of publication bias in research on psychological interventions is indirect and the result of an extrapolation of observed small study effects described in the previous section. For instance, for trials of psychological therapies compared to control conditions for adult depression, Cuijpers, Smit, Bohlmeijer, Hollon, and Andersson (2010) used a series of tests to document strong and significant indication of small study effects, which they attributed to publication bias. Analyses restricted to trials of CBT, which represented the largest share of the data, pointed to a similar pattern. Another possible proxy for publication bias is the test of excess significance, which determines whether the number of statistically significant results in a pool of studies is too high, based on some plausible assumptions about the magnitude of the true effect (Ioannidis & Trikalinos, 2007). Using this method, an excess of significant findings was reported for trials of psychotherapy for depression, and the pattern was maintained in sensitivity analyses restricted to trials with a CBT arm (Flint, Cuijpers, Horder, Koole, & Munafò, 2015). More broadly, an umbrella review of meta-analyses of psychotherapy for various symptom, distress and physical outcomes reported excess significance in around 40% of the included meta-analyses (Dragioti et al., 2017). However, another umbrella review of psychological interventions for pain found very limited evidence of excess significance, in just 7% of the included meta-analyses (Markozannes et al., 2017).

Nevertheless, both small study effects and excess significance offer only indirect and speculative evidence toward publication bias and have their own limitations. Though publication bias is often the most plausible cause, small study effects can also result from other factors, such as clinical heterogeneity between patients in large and small trials, a mathematical artefact or mere coincidence (Rücker, Carpenter, & Schwarzer, 2011). The excess significance test relies on assumptions about the magnitude of the true effect, which in practice is often equated with the effect from the largest study. However, there are few large trials of psychological interventions. Moreover, a large study can have design, implementation or reporting flaws, and hence be at high risk of bias. The summary effect of a meta-analysis is another possible estimation of the true effect; yet, if there is indeed excess significance, it too would be affected.

Publication bias can only be directly and unequivocally determined if access is ensured to all conducted studies, for example all trials submitted to a regulatory authority. In a seminal report using the Food and Drug Administration (FDA) database, E. H. Turner, Matthews,

Linardatos, Tell, and Rosenthal (2008) showed that among 74 antidepressant trials, 31% remained unpublished. Separate meta-analyses of the FDA and journal datasets showed publication bias accounted for a relative effect size increase of 32%. Thirty-seven out of the 38 studies characterized by the FDA as having positive results were published. In contrast, except for 3 studies, trials viewed by the FDA as having negative or questionable results were mostly not published (22 studies) or published in a way that conveyed a positive outcome (11 studies).

An authority similar to the FDA does not exist for psychological interventions, making direct assessments of publication bias more complex and often unfeasible. In one of the rare evaluations, Driessen, Hollon, Bockting, Cuijpers, and Turner (2015) identified US National Institutes of Health grants awarded for randomized trials of psychotherapy for depression and matched grants with published reports. For grants that had not resulted in publications, data was requested from investigators and both published and unpublished findings were meta-analysed. About a quarter of the publicly funded grants did not result in publications, an estimate of the extent of publication bias. Unpublished studies resulted in a small and non-significant effect size (Hedges' g) of 0.20, and their addition to the published studies reduced the summary effect by 25%, to $g = 0.39$. Notably, this investigation closely parallels previously described findings regarding the selective publication of antidepressant trials (E. H. Turner et al., 2008).

The landmark work of E. H. Turner et al. (2008) also demonstrates how the conclusions of published papers can deviate substantially from what the data really show. Yet without access to datasets, deviations are impossible to detect. An example perhaps even more dramatic example of how lack of access to the complete trial data can entirely alter the interpretation of the findings is provided by the re-analysis of SmithKline Beecham's Study 329, comparing paroxetine, imipramine and placebo for adolescent unipolar depression (Le Noury et al., 2015). This re-analysis relied on clinical study reports previously not publicly available and showed that none of the drugs outperformed placebo, but both were also associated with significant increases in serious harms, starkly contradicting the published report (Keller et al., 2001). Inaccessibility of complete individual patient data (IPD) can thus have catastrophic consequences.

Therefore, IPD underlying reported findings are another crucial class of information. Data sharing could potentially fulfil a multitude of key functions, with far-reaching implications for patients, clinicians, researchers and other users of scientific literature. It would presumably contribute to resolving the above-mentioned flaws in the current system of communicating research results, such as selective reporting, distortion of evidence and lack of reproducibility. In addition, research value increases with greater transparency and the opportunity for external researchers to re-analyse (Ebrahim et al., 2014; Naudet et al., 2018), synthesize (e.g., IPD meta-analyses) (Tannenbaum et al., 2018), or build upon previous data (including secondary analyses and methodological work). To take just one example: our own commentary (Cristea et al., 2018) of the Iyadurai et al. (2018) Tetris trial was possible because the authors shared their data, thus allowing for an independent analysis and, based on it, a divergent interpretation. Not only does this practice stimulate vigorous scientific debate, a cornerstone of scientific progress, but it also prevents the waste of collecting new data to explore the same question.

Yet, for all of its expected benefits and despite recent recommendations from the International Committee of Medical Journal Editors (ICMJE, Clinical Trials- Data sharing, 2019), the adoption of data sharing practices remains very low, by both authors and journals. Cross-sectional analyses of the biomedical literature showed derisory rates of data sharing, ranging from 0 (Iqbal, Wallach, Khoury, Schully, & Ioannidis, 2016) to 11% (Wallach, Boyack, & Ioannidis, 2018). For clinical psychology specifically, an audit of the highest-ranking journals (Nutu, Gentili, Naudet, & Cristea, 2019) found that even if two thirds of the journals recommended data sharing, only one had a mandatory policy in place. Compliance with the recommended policies in a sample

of recently published articles was extremely low (2%), in line with reports from the biomedical literature. Of course, barriers need to be acknowledged, and a systematic review of data sharing in public health (van Panhuis et al., 2014) identified several real or potential ones, ranging from technical (e.g., lack of availability of technical solutions and of metadata and standards), motivational (e.g., lack of incentives, disagreement on data use), economic (e.g., lack of resources), political (e.g., lack of trust or guidelines), legal (e.g., ownership and copyright, protection of privacy) and ethical (e.g., lack of reciprocity). An important element in overcoming barriers is crediting and rewarding investigators for sharing data, along with other artefacts such as materials or code. Citation is one basic form of awarding credit (Alter & Gonzalez, 2018) and datasets can become citable by receiving a persistent identifier, such as a digital object identifier (DOI), through platforms like DataCite (<https://datacite.org/>). Looking forward, owing to its conspicuous societal value, sharing of data (but also of protocols, materials, code etc) could be rewarded and incentivized, by incorporating it in assessments of scientists for hiring, promotion and tenure (Moher et al., 2018). Another important issue is that trials with small samples, frequent in psychotherapy research, pose a higher risk of partial de-identification, particularly for under-represented patient categories. In other cases, trials have been underway for a long time, and consent forms might have not explicitly asked about sharing data. Yet, it is important to underscore that patients do not seem to share these worries: only 8% of clinical trial participants in a recent survey (Mello, Lieou, & Goodman, 2018) felt that potentially harmful consequences of data sharing outweighed the benefits and around 26% were concerned about being identified.

4. Unbiased and useable research reports?

Research reports should clearly and transparently delineate “*what questions were addressed and why, what was done, what was shown and what the findings mean*” (Glasziou et al., 2014) (p.267). Reporting guidelines, which originated with the Consolidated Standards for Reporting Trials (CONSORT) statement (Schulz, Altman, & Moher, 2010), are one effective way to ensure this happens. These offer guidance and structured forms on how to report research methods and findings (Blanco et al., 2017), specific to different study types (“*Library for health research reporting*”). The EQUATOR Network (<https://www.equator-network.org/about-us/>), a portal dedicated to augmenting research value by promoting the use of reporting guidelines, lists over 400 of these, tailored for a diversity of research designs, populations and interventions. Though their consistent utilization would reduce the danger of incomplete, misleading or otherwise unusable reports, adoption from journals and compliance from researchers are both still wanting. An early systematic review (Grant, Mayo-Wilson, Melendez-Torres, & Montgomery, 2013), which surveyed the adoption of reporting guidelines in high-impact journals in several fields publishing research on psychosocial interventions, found that around a quarter of the surveyed journals mentioned guidelines in their instructions to authors. For a sample of trials published in these journals, compliance to reporting standards was around 42%. In a more recent audit of top-tier journals in clinical psychology (Nutu, Gentili, Naudet, & Cristea, 2019), approximately half of the surveyed journals explicitly endorsed reporting guidelines in their instructions to author. However, the more stringent requirement of making use of guidelines mandatory was in place at 30% of the journals for randomized controlled trials, and at 25% of them for systematic reviews and meta-analyses. An analysis of a cross-sectional sample of articles recently published in the surveyed journals showed breaches in adoption still remained, even when guidelines were mandatory as per journal policy: 3/14 randomized trials and 7/24 systematic reviews and meta-analyses were rated as non-compliant.

Glasziou et al. (2014) warn that waste is not restricted to study reports. In order to be able to judge what was done, it is essential to

know what was planned. Hence, it is also critical that the study protocol and statistical analysis plan are unbiased and useable. Complete and transparent prospective registration of the study plan in a public repository and the publication of the study protocol preceding its implementation are potential ways to ensure this. These measures would also be key in minimizing bias due to selective outcome reporting, hence limiting research waste due to inappropriate methodological choices, described in the previous section. Unfortunately, prospective registration is still underutilized in research on psychological interventions. Meta-epidemiological studies show low rates of compliance: 25% in a survey of 25 highest-ranking clinical psychology journals that published randomized controlled trials (Cybulski, Mayo-Wilson, & Grant, 2016), and 50% in one of trials published in 2013–2014 in the *Journal of Consulting and Clinical Psychology*, the American Psychological Association’s primary outlet for this study type (Azar, Riehm, McKay, & Thombs, 2015). To map the expanse of the problem, we also looked upstream at what journals are requiring from their authors (Nutu, Gentili, Naudet, & Cristea, 2019). Under 40% of the highest-ranking journals in clinical psychology mentioned registration in their instructions to authors, and only 25% specified it needed to be prospective (i.e., before the enrolment of the 1st participant). Compliance with registration requirements in a sample of articles published in these journals in 2017 was extremely low: 3% if only prospective registration was considered, and 8% for both prospective and retrospective registration.

Finally, another source of waste is fragmentation, which can take several guises: 1) “salami-slicing”, partitioning findings from the same study into a number of thin “slices” or least publishable units (Martin, 2013); 2) reporting results across a multiplicity of sources (e.g., journal articles, trial registries, clinical study reports) (Li, Mayo-Wilson, Fusco, Hong, & Dickersin, 2018); or 3) running many parallel non-collaborative efforts. Though we could not identify investigations of fragmentation and multiplicity in psychological interventions research, an analysis of trials for the drugs gabapentin for neuropathic pain and quetiapine for bipolar depression showed that information about trial design and risk of bias frequently diverged in completeness across reports. Different reports of the same data often conflicted, in ways that were difficult to disentangle (Mayo-Wilson et al., 2017). A particularly serious problem occurs when fragmentation of results also comprises non-public sources, such as clinical study reports. The re-analysis of SmithKline Beecham’s Study 329 (Le Noury et al., 2015) is a revealing example, because it was based on clinical study reports previously not publicly accessible. Multiplicity also enhances the hazardous effects of selective reporting beyond single trials to meta-analyses, by allowing researchers to cherry-pick outcomes and results and conduct many “possible” and even conflicting meta-analyses (Mayo-Wilson et al., 2017).

Overlapping meta-analyses on the same topic are a good example of waste due to fragmentation of efforts. Meta-analyses are expected to reduce waste by helping delineate research questions worth exploring and by delivering a reproducible and quantitative synthesis on a given question. However, their number has raised to epidemic proportions in the past decade, possibly because they amount to easily produced publishable units or powerful marketing tools (Ebrahim, Bance, Athale, Malachowski, & Ioannidis, 2016). For instance, the database of trials of psychotherapy for depression (Cuijpers, van Straten, Warmerdam, & Andersson, 2008) (<http://www.evidencebasedpsychotherapies.org/>) included, at the last publicly available update to the website in 2014, 352 trials and 53 published meta-analyses of these, a rate of approximately 1 meta-analysis for every 7 trials. Though widely viewed as a higher level of evidence and as the authoritative final word in a debate, meta-analyses are often suboptimal, overlapping and sometimes in contradiction (Ioannidis, 2016a). Although the basic principles of a meta-analysis are simple, researchers can make many choices along the way that can significantly alter the results and sometimes lead to opposing conclusions. In an egregious simulation, Palpacuer et al. (2019)

demonstrated that varying methodological criteria and analytical models resulted into 9216 possible different and contradictory meta-analyses of indirect comparisons, for 60 primary trials of nalmefene and naltrexone in alcohol disorders.

Similarly, among actually published meta-analyses, the effectiveness of long-term psychodynamic psychotherapy for mental disorders was estimated in one meta-analysis with a risk difference of 0 for the primary outcome of recovery at the longest follow-up and with non-significant effect sizes for other outcomes (Smit et al., 2012), while in another meta-analysis (Leichsenring & Rabung, 2011) the same treatment produced significant and moderate effect sizes (Hedges' g) for overall effectiveness, as well as for other outcomes. In another example, two meta-analyses of acceptance and commitment therapy (ACT) resulted into largely similar effect sizes for psychiatric and somatic disorders (differences in Hedges' g between the two meta-analyses ranged from 0.10 to 0.15), but reached widely divergent conclusions: supporting the use of the therapy in one (A. Tjak et al., 2015), and questioning its status as a "well-established" treatment in another (Ost, 2014). In both examples, meta-analyses published within a year of each other diverged in the magnitude or even significance of the effect, and, even more concerningly, supported markedly discrepant conclusions about treatment effectiveness. Though IPD meta-analysis are generally seen as powerful tools in identifying individual-level variables that could act as effect modifiers (Riley, Lambert, & Abo-Zaid, 2010), only a small fraction (under 10%) of them uncover statistically significant interactions for treatment effects, translating into a modest potential to personalize treatment (Schuit, Li, & Ioannidis, 2018).

5. Solutions for increasing value

We used the Lancet series on reducing waste and increasing value as a template to discuss the state of research in the field of clinical psychology with a particular focus on psychological therapies. The literature we reviewed demonstrates that all sources of waste considered are well-represented and amply documented within this field. We did not discuss waste related to research regulation and management (Salman et al., 2014), because, in contrast to biomedicine, research on psychological treatments is neither formally regulated, nor does it have to comply with the requirements of regulatory bodies like the FDA or the European Medicine Agency. Our goal was to exemplify varieties of waste and as such our analysis was neither systematic, nor exhaustive. Unlike Chalmers and Glasziou (2009), we did not attempt to quantify the proportion of avoidable waste from research investment. Whether the 85% estimated for biomedicine is an under- or an overestimation remains an open question. Still, we hope our evaluation can serve as a warning call to all those involved with psychological treatments, from funders to researchers and patients, to stimulate reflection and, when possible, reforms to increase value in research on psychological interventions.

In response to the problems diagnosed and extending the solutions suggested in the Lancet series, several measures were proposed to optimize key elements of the scientific process, such as methods, reporting and dissemination, reproducibility, evaluation and initiatives (Munafò et al., 2017). A discussion of these is beyond the scope of this paper, but most, if not all, could be successfully tested with the goal of improving research on psychological interventions. Similarly, Tackett et al. (2017) scrutinized possible reasons why the conversation around reproducibility has remained peripheral in clinical psychology, such as a greater appreciation of complex and imperfect data or more variable sampling procedures. The authors also suggest challenges and recommendations in reference to reproducibility concerns specific to clinical psychology.

We further expand on a few general suggestions. First, the emerging field of meta-research ("research on research") is essential for monitoring and diagnosing the "health" of psychological treatment research. Though science is often equated with primary research, inadequately planned, conducted or reported primary research, described

in biased or unusable reports, with findings or data not fully accessible, is at best inconsequential and at worst harmful to patients and others studying or employing psychological treatments. Secondary/meta-research is indispensable in surveilling the integrity and transparency of the research process and in detecting systematic biases. This field includes research synthesis, meta-analysis and various other types of meta-epidemiological assessment, which critically examine the ways in which studies are planned, conducted or reported, or that seek to detect the influences of systematic biases in large collections of studies. Funders, academic institutions and journal editors need to recognize the role of and encourage meta-research. Yet many meta-research investigations themselves suffer from the same waste-generating issues we identified, such as duplicating what is already known or researched, not being prospectively registered, or not sharing their underlying data. Interventions intended to improve research quality should be planned and tested in adequately conducted and reported randomized trials (Bravo, Grimaldo, Lopez-Inesta, Mehmani, & Squazzoni, 2019); otherwise, they risk adding to the waste. For instance, only a handful of interventions to improve the quality of peer-review have been assessed in randomized trials, with mixed results (Bruce, Chauvin, Trinquart, Ravaud, & Boutron, 2016).

Furthermore, open science principles, such as transparency, ensuring that studies are accurately and fully described across all stages, from planning to the interpretation of findings, accessibility of data underlying results, reproducibility checks and independent replication at the very least of findings lined up for clinical translation, need to be placed at the core of psychological treatment research. It was somewhat surprising that the programmatic Lancet Psychiatry Commission on psychological treatments (Holmes et al., 2018) did not attach greater importance and a more central part to open science (Cristea, 2018a). As we amply documented in this comment, deviations from open science principles and practices underlie most failures to engender transformative and useful research. Funders, journal editors and academic institutions could test the usefulness of the Transparency and Openness Promotion (TOP) guidelines (Nosek et al., 2015) to incentivize the principles of transparency, openness and reproducibility in planning, conducting and reporting research.

Finally, we need to remain vigilant particularly of the strains of research currently prioritized, such as experimental psychopathology. The hype that often accompanies findings in this field is most likely detrimental and should be contained. The fact that a particular result could have major treatments implications should not distract us from ensuring that the methods by which it was obtained are sound, its reporting complete and transparent, and independent replication does occur. The hypothetical promise of a particular type of theory or model cannot supplant modest or fleeting effects, which are unlikely to engender any tangible benefit to patients. Furthermore, if our aim is for psychological interventions to be developed mechanistically, following a process similar to the development and testing of drugs, it must also be acknowledged that, as with most initially encouraging drug targets, many originally promising hypotheses stemming from the laboratory will need to be abandoned when they fail replication, negative findings accumulate, or when independent, well-conducted meta-analyses show small and unreliable effects, lack of effects in clinical populations, or critical issues in the planning, conducting and reporting of studies. Even if all these conditions were met, translation from proof-of-concept trials might result in weak, unstable or non-durable effects which might not justify the immense investment of a phase III trial.

In a thought-provoking comment, Ioannidis (2016b) argued that since most psychological interventions probably do not work, they should be examined in "carefully properly biased" (p.437) studies, where investigators give their best shot to augment the effectiveness of the intervention, by increasing expectations, boosting any placebo effects, using waiting list control arms and small samples. The phase would function as a pre-screening and only the interventions attaining a reasonable effect under these very favorable conditions would move to the

phase of testing in a larger trial. While appealing, this system could never work if other sources of waste are not tackled: studies poorly planned and implemented, reports unusable and incomplete, negative findings stashed away, or data not accessible. Ultimately, if research and meta-research on psychological interventions aims to become more mechanistic and science-based, it should heed Douglas Altman's famous proclamation "We need less research, better research, and research done for the right reasons" (Altman, 1994) (p.308).

Acknowledgements

Financial support

Ioana A. Cristea was supported by a grant from the Romanian Ministry of Research and Innovation, CNCS – UEFISCDI (project number PN-III-P1-1.1-TE-2016-1054). The funder had no role in the design of the study, collection, analysis and interpretation of data, and the decision to approve publication.

We would like to thank Dr. Tom E. Hardwicke for comments on an early draft of the manuscript.

Conflict of interest disclosure

All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coi_disclosure.pdf (available upon request from the corresponding author) and declare no conflict of interest.

References

- Alter, G., & Gonzalez, R. (2018). Responsible practices for data sharing. *American Psychologist*, 73, 146–156.
- Altman, D. G. (1994). The scandal of poor medical research. *BMJ*, 308, 283–284.
- Anderson, N. L. (2010). The clinical plasma proteome: A survey of clinical assays for proteins in plasma and serum. *Clinical Chemistry*, 56, 177–185.
- Azar, M., Riehm, K. E., McKay, D., & Thombs, B. D. (2015). Transparency of outcome reporting and trial registration of randomized controlled trials published in the journal of consulting and clinical psychology. *PLoS One*, 10, e0142894.
- Barth, J., Munder, T., Gerger, H., Nuesch, E., Trelle, S., Znoj, H., et al. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: A network meta-analysis. *PLoS Medicine*, 10, e1001454.
- Bighelli, I., Salanti, G., Huhn, M., Schneider-Thoma, J., Krause, M., Reitmair, C., et al. (2018). Psychological interventions to reduce positive symptoms in schizophrenia: Systematic review and network meta-analysis. *World Psychiatry*, 17, 316–329.
- Bisson, J. I., Roberts, N. P., Andrew, M., Cooper, R., & Lewis, C. (2013). *Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults*. Cochrane Database Syst Rev Cd003388.
- Blanco, D., Kirkham, J. J., Altman, D. G., Moher, D., Boutron, I., & Cobo, E. (2017). Interventions to improve adherence to reporting guidelines in health research: A scoping review protocol. *BMJ Open*, 7.
- Bradley, H. A., Rucklidge, J. J., & Mulder, R. T. (2017). A systematic review of trial registration and selective outcome reporting in psychotherapy randomized controlled trials. *Acta Psychiatrica Scandinavica*, 135, 65–77.
- Bravo, G., Grimaldo, F., Lopez-Inesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications*, 10, 322.
- Browning, M., Holmes, E. A., Charles, M., Cowen, P. J., & Harmer, C. J. (2012). Using attentional bias modification as a cognitive vaccine against depression. *Biological Psychiatry*, 72, 572–579.
- Bruce, R., Chauvin, A., Trinquart, L., Ravaut, P., & Boutron, I. (2016). Impact of interventions to improve the quality of peer review of biomedical journals: A systematic review and meta-analysis. *BMC Medicine*, 14, 85.
- Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A. M., et al. (2014). How to increase value and reduce waste when research priorities are set. *The Lancet*, 383, 156–165.
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, 374, 86–89.
- Chan, A.-W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P. C., et al. (2014). Increasing value and reducing waste: Addressing inaccessible research. *The Lancet*, 383, 257–266.
- Chatterton, M. L., Stockings, E., Berk, M., Barendregt, J. J., Carter, R., & Mihalopoulos, C. (2017). Psychosocial therapies for the adjunctive treatment of bipolar disorder in adults: Network meta-analysis. *British Journal of Psychiatry*, 210, 333–341.
- Contopoulos-Ioannidis, D. G., Ntzani, E. E., & Ioannidis, J. P. A. (2003). Translation of highly promising basic science research into clinical applications. *The American Journal of Medicine*, 114, 477–484.
- Cristea, I. A. (2018a). Experimental psychopathology should not be examined with rose-tinted glasses. *The Lancet Psychiatry*, 5, 465–466.
- Cristea, I. A. (2018b). The waiting list is an inadequate benchmark for estimating the effectiveness of psychotherapy for depression. *Epidemiology and Psychiatric Sciences*, 1–2.
- Cristea, I. A., Gentili, C., Cotet, C. D., Palomba, D., Barbui, C., & Cuijpers, P. (2017). Efficacy of psychotherapies for borderline personality disorder: A systematic review and meta-analysis. *JAMA Psychiatry*, 74, 319–328.
- Cristea, I. A., Gentili, C., Pietrini, P., & Cuijpers, P. (2017a). Is investigator background related to outcome in head to head trials of psychotherapy and pharmacotherapy for adult depression? A systematic review and meta-analysis. *PLoS One*, 12, e0171654.
- Cristea, I. A., Gentili, C., Pietrini, P., & Cuijpers, P. (2017b). Sponsorship bias in the comparative efficacy of psychotherapy and pharmacotherapy for adult depression: Meta-analysis. *British Journal of Psychiatry*, 210, 16–23.
- Cristea, I. A., & Ioannidis, J. P. A. (2018). Improving disclosure of financial conflicts of interest for research on psychosocial interventions. *JAMA Psychiatry*, 75, 541–542.
- Cristea, I. A., Kok, R. N., & Cuijpers, P. (2015). Efficacy of cognitive bias modification interventions in anxiety and depression: meta-analysis. *British Journal of Psychiatry*, 206, 7–16.
- Cristea, I. A., Mogoase, C., David, D., & Cuijpers, P. (2015). Practitioner review: Cognitive bias modification for mental health problems in children and adolescents: A meta-analysis. *Journal of Child Psychology and Psychiatry*, 56, 723–734.
- Cristea, I. A., Naudet, F., Shanks, D. R., & Hardwicke, T. E. (2018). Post-retrieval Tetris should not be likened to a 'cognitive vaccine'. *Molecular Psychiatry*, 23, 1972–1973.
- Crocker, J. C., Ricci-Cabello, I., Parker, A., Hirst, J. A., Chant, A., Petit-Zeman, S., et al. (2018). Impact of patient and public involvement on enrolment and retention in clinical trials: Systematic review and meta-analysis. *BMJ*, 363, k4738.
- Cuijpers, P. (2019). Targets and outcomes of psychotherapies for mental disorders: An overview. *World Psychiatry*, 18(3), 276–285. <https://doi.org/10.1002/wps.20661>.
- Cuijpers, P., & Cristea, I. A. (2016). How to prove that your therapy is effective, even when it is not: A guideline. *Epidemiology and Psychiatric Sciences*, 25, 428–435.
- Cuijpers, P., Cristea, I. A., Karyotaki, E., Reijnders, M., & Huibers, M. J. (2016). How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry*, 15, 245–258.
- Cuijpers, P., Donker, T., Weissman, M. M., Ravitz, P., & Cristea, I. A. (2016). Interpersonal psychotherapy for mental health problems: A comprehensive meta-analysis. *American Journal of Psychiatry*, 173, 680–687.
- Cuijpers, P., Karyotaki, E., Reijnders, M., & Ebert, D. D. (2019). Was eyesenck right after all? A reassessment of the effects of psychotherapy for adult depression. *Epidemiology and Psychiatric Sciences*, 28, 21–30.
- Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias. *British Journal of Psychiatry*, 196, 173–178.
- Cuijpers, P., van Straten, A., Warmerdam, L., & Andersson, G. (2008). Psychological treatment of depression: A meta-analytic database of randomized studies. *BMC Psychiatry*, 8, 36.
- Cybulski, L., Mayo-Wilson, E., & Grant, S. (2016). Improving transparency and reproducibility through registration: The status of intervention trials published in clinical psychology journals. *Journal of Consulting and Clinical Psychology*, 84, 753–767.
- Deighton, J., Lereya, S. T., Casey, P., Patalay, P., Humphrey, N., & Wolpert, M. (2019). Prevalence of mental health problems in schools: Poverty and other risk factors among 28 000 adolescents in England. *British Journal of Psychiatry*, 1–3.
- Dragioti, E., Dimoliatis, I., & Evangelou, E. (2015). Disclosure of researcher allegiance in meta-analyses and randomised controlled trials of psychotherapy: A systematic appraisal. *BMJ Open*, 5, e007206.
- Dragioti, E., Dimoliatis, I., Fountoulakis, K. N., & Evangelou, E. (2015). A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Annals of General Psychiatry*, 14, 25.
- Dragioti, E., Karathanos, V., Gerdle, B., & Evangelou, E. (2017). Does psychotherapy work? An umbrella review of meta-analyses of randomized controlled trials. *Acta Psychiatrica Scandinavica*, 136, 236–246.
- Drissen, E., Hollon, S. D., Bockting, C. L., Cuijpers, P., & Turner, E. H. (2015). Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US national Institutes of health-funded trials. *PLoS One*, 10, e0137864.
- Ebrahim, S., Bance, S., Athale, A., Malachowski, C., & Ioannidis, J. P. (2016). Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *Journal of Clinical Epidemiology*, 70, 155–163.
- Ebrahim, S., Sohani, Z. N., Montoya, L., Agarwal, A., Thorlund, K., Mills, E. J., et al. (2014). Reanalyses of randomized clinical trial data. *Jama*, 312, 1024–1032.
- Flint, J., Cuijpers, P., Horder, J., Koole, S. L., & Munafò, M. R. (2015). Is there an excess of significant findings in published studies of psychotherapy for depression? *Psychological Medicine*, 45, 439–446.
- Fodor, L. A., Cotet, C. D., Cuijpers, P., Szamoskozi, S., David, D., & Cristea, I. A. (2018). The effectiveness of virtual reality based interventions for symptoms of anxiety and depression: A meta-analysis. *Scientific Reports*, 8, 10323.
- Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., et al. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, 130, 181–192.
- Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., et al. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, 383, 267–276.
- Grant, S. P., Mayo-Wilson, E., Melendez-Torres, G. J., & Montgomery, P. (2013). Reporting quality of social and psychological intervention trials: A systematic review of reporting guidelines and trial publications. *PLoS One*, 8, e65442.
- Hardwicke, T. E., & Shanks, D. R. (2016). Reply to Walker and Stickgold: Proposed

- boundary conditions on memory reconsolidation will require empirical verification. *Proceedings of the National Academy of Sciences of the United States of America*, 113, E3993–E3994.
- Hardwicke, T. E., Taqi, M., & Shanks, D. R. (2016). Postretrieval new learning does not reliably induce human memory updating via reconsolidation. *Proceedings of the National Academy of Sciences of the United States of America*, 113, 5206–5211.
- Hazo, J. B., Gandre, C., Leboyer, M., Obradors-Tarrago, C., Belli, S., McDaid, D., et al. (2017). National funding for mental health research in Finland, France, Spain and the United Kingdom. *European Neuropsychopharmacology*, 27, 892–899.
- Hazo, J. B., Gervais, J., Gandre, C., Brunn, M., Leboyer, M., & Chevreur, K. (2016). European Union investment and countries' involvement in mental health research between 2007 and 2013. *Acta Psychiatrica Scandinavica*, 134, 138–149.
- Higgins, J. P. T., Altman, D. G., Göttsche, P. C., Jüni, P., Moher, D., Oxman, A. D., et al. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*, 343, d5928.
- Higgins, J. P. T., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions*. The Cochrane Collaboration Version 5.1.0. ed.
- Holmes, E. A., Ghaderi, A., Eriksson, E., Lauri, K. O., Kukacka, O. M., Mamish, M., et al. (2017). 'I can't concentrate': A feasibility study with young refugees in Sweden on developing science-driven interventions for intrusive memories related to trauma. *Behavioural and Cognitive Psychotherapy*, 45, 97–109.
- Holmes, E. A., Ghaderi, A., Harmer, C. J., Ramchandani, P. G., Cuijpers, P., Morrison, A. P., et al. (2018). The *Lancet Psychiatry* Commission on psychological treatments research in tomorrow's science. *The Lancet Psychiatry*, 5, 237–286.
- Holmes, E. A., James, E. L., Kilford, E. J., & Deeprose, C. (2010). Key steps in developing a cognitive vaccine against traumatic flashbacks: Visuospatial Tetris versus verbal pub quiz. *PLoS One*, 5, e13706.
- ICMJE. *Clinical Trials- Data sharing*. In (Vol. 2019).
- Ioannidis, J. P. A. (2016a). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, 94, 485–514.
- Ioannidis, J. P. A. (2016b). Most psychotherapies do not really work, but those that might work should be assessed in biased studies. *Epidemiology and Psychiatric Sciences*, 25, 436–438.
- Ioannidis, J. P. A., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., et al. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*, 383, 166–175.
- Ioannidis, J. P. A., & Trikalinos, T. A. (2007). An exploratory test for an excess of significant findings. *Clinical Trials*, 4, 245–253.
- Iqbal, S. A., Wallach, J. D., Khoury, M. J., Schully, S. D., & Ioannidis, J. P. A. (2016). Reproducible research practices and transparency across the biomedical literature. *PLoS Biology*, 14, e1002333.
- Iyadurai, L., Blackwell, S. E., Meiser-Stedman, R., Watson, P. C., Bonsall, M. B., Geddes, J. R., et al. (2018). Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: A proof-of-concept randomized controlled trial. *Molecular Psychiatry*, 23, 674–682.
- Jayasuriya, D., Jayasuriya, R., Tay, A. K., & Silove, D. (2016). Associations of mental distress with residency in conflict zones, ethnic minority status, and potentially modifiable social factors following conflict in Sri Lanka: A nationwide cross-sectional study. *The Lancet Psychiatry*, 3, 145–153.
- Keller, M. B., Ryan, N. D., Strober, M., Klein, R. G., Kutcher, S. P., Birmaher, B., et al. (2001). Efficacy of paroxetine in the treatment of adolescent major depression: A randomized, controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 40, 762–772.
- Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W. A. (2012). A systematic review of comparative efficacy of treatments and controls for depression. *PLoS One*, 7, e41778.
- Kindt, M., & Soeter, M. (2013). Reconsolidation in a human fear conditioning study: A test of extinction as updating mechanism. *Biological Psychology*, 92, 43–50.
- Kruijt, A.-W., Parsons, S., & Fox, E. (2019). A meta-analysis of bias at baseline in RCTs of attention bias modification: No evidence for dot-probe bias towards threat in clinical anxiety and PTSD. *Journal of Abnormal Psychology*, 128(6), 563–573. <http://dx.doi.org/10.1037/abn0000406>.
- Kyzas, P. A., Denaxa-Kyza, D., & Ioannidis, J. P. A. (2007). Almost all articles on cancer prognostic markers report statistically significant results. *European Journal of Cancer*, 43, 2559–2579.
- Le Noury, J., Nardo, J. M., Healy, D., Jureidini, J., Raven, M., Tufanaru, C., et al. (2015). Restoring study 329: Efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ*, 351, h4320.
- Leichsenring, F., & Rabung, S. (2011). Long-term psychodynamic psychotherapy in complex mental disorders: Update of a meta-analysis. *British Journal of Psychiatry*, 199, 15–22.
- Library for health research reporting EQUATOR Network. <http://www.equator-network.org/reporting-guidelines/>, Accessed date: 24 July 2019.
- Li, T., Mayo-Wilson, E., Fusco, N., Hong, H., & Dickersin, K. (2018). Caveat emptor: The combined effects of multiplicity and selective reporting. *Trials*, 19, 497.
- Lovell, K., Bower, P., Gellatly, J., Byford, S., Bee, P., McMillan, D., et al. (2017). Low-intensity cognitive-behaviour therapy interventions for obsessive-compulsive disorder compared to waiting list for therapist-led cognitive-behaviour therapy: 3-arm randomised controlled trial of clinical effectiveness. *PLoS Medicine*, 14, e1002337.
- MacLeod, C. (2012). Cognitive bias modification procedures in the management of mental disorders. *Current Opinion in Psychiatry*, 25, 114–120.
- MacLeod, C., & Mathews, A. (2012). Cognitive bias modification approaches to anxiety. *Annual Review of Clinical Psychology*, 8, 189–217.
- Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P. A., et al. (2014). Biomedical research: Increasing value, reducing waste. *The Lancet*, 383, 101–104.
- Markozannes, G., Aretouli, E., Rintou, E., Dragioti, E., Damigos, D., Ntzani, E., et al. (2017). An umbrella review of the literature on the effectiveness of psychological interventions for pain reduction. *BMC Psychol*, 5, 31.
- Martin, B. R. (2013). Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment. *Research Policy*, 42, 1005–1014.
- Mayo-Wilson, E., Dias, S., Mavranzeouli, I., Kew, K., Clark, D. M., Ades, A. E., et al. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 1, 368–376.
- Mayo-Wilson, E., Li, T., Fusco, N., Bertizzolo, L., Canner, J. K., Cowley, T., et al. (2017). Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *Journal of Clinical Epidemiology*, 91, 95–110.
- Meichenbaum, D., & Lilienfeld, S. O. (2018). How to spot hype in the field of psychotherapy: A 19-item checklist. *Professional Psychology: Research and Practice*, 49, 22–30.
- Mello, M. M., Lieou, V., & Goodman, S. N. (2018). Clinical trial participants' views of the risks and benefits of data sharing. *New England Journal of Medicine*, 378, 2202–2211.
- Moher, D., Naudet, F., Cristea, I. A., Miedema, F., Ioannidis, J. P. A., & Goodman, S. N. (2018). Assessing scientists for hiring, promotion, and tenure. *PLoS Biology*, 16, e2004089.
- MQ (2019). UK mental health research funding 2014–2017. *MQ: Transforming mental health*.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., et al. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- Munder, T., Fluckiger, C., Gerger, H., Wampold, B. E., & Barth, J. (2012). Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? A meta-analysis. *Journal of Counseling Psychology*, 59, 631–637.
- Munder, T., Gerger, H., Trelle, S., & Barth, J. (2011). Testing the allegiance bias hypothesis: A meta-analysis. *Psychotherapy Research*, 21, 670–684.
- Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, 406, 722–726.
- Naudet, F., Sakarovich, C., Janiaud, P., Cristea, I., Fanelli, D., Moher, D., et al. (2018). Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in The BMJ and PLOS medicine . *BMJ*, 360.
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science*, 348, 1422–1425.
- Nutu, D., Gentili, C., Naudet, F., & Cristea, I. A. (2019). Open science practices in clinical psychology journals: An audit study. *Journal of Abnormal Psychology*, 128(6), 510–516. <http://dx.doi.org/10.1037/abn0000414>.
- Ost, L. G. (2014). The efficacy of acceptance and commitment therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, 61, 105–121.
- Palpacuer, C., Hammam, K., Duprez, R., Laviolle, B., Ioannidis, J. P. A. I., & Naudet, F. (under review). Vibration of effects from diverse inclusion/exclusion criteria and analytical choices: 9,216 different ways to perform an indirect comparison meta-analysis. *BMC Medicine*.
- van Panhuis, W. G., Paul, P., Emerson, C., Grefenstette, J., Wilder, R., Herbst, A. J., et al. (2014). A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14, 1144.
- Patel, V., Burns, J. K., Dhingra, M., Tarver, L., Kohrt, B. A., & Lund, C. (2018). Income inequality and depression: A systematic review and meta-analysis of the association and a scoping review of mechanisms. *World Psychiatry*, 17, 76–89.
- Pompoli, A., Furukawa, T. A., Imai, H., Tajika, A., Fthimiou, O., & Salanti, G. (2016). Psychological therapies for panic disorder with or without agoraphobia in adults: A network meta-analysis. *Cochrane Database of Systematic Reviews*, 13(4), <https://doi.org/10.1002/14651858.CD011004.pub2>.
- Riley, R. D., Lambert, P. C., & Abo-Zaid, G. (2010). Meta-analysis of individual participant data: Rationale, conduct, and reporting. *BMJ*, 340, c221.
- Rücker, G., Carpenter, J. R., & Schwarzer, G. (2011). Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal*, 53, 351–368.
- Salman, R. A.-S., Beller, E., Kagan, J., Hemminki, E., Phillips, R. S., Savulescu, J., et al. (2014). Increasing value and reducing waste in biomedical research regulation and management. *The Lancet*, 383, 176–185.
- Schiller, D., Monfils, M.-H., Raio, C. M., Johnson, D. C., LeDoux, J. E., & Phelps, E. A. (2009). Preventing the return of fear in humans using reconsolidation update mechanisms. *Nature*, 463, 49.
- Schuit, E., Li, A. H., & Ioannidis, J. P. A. (2018). How often can meta-analyses of individual-level data individualize treatment? A meta-epidemiologic study. *International Journal of Epidemiology*, 48, 596–608.
- Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, 152, 726–732.
- Shinohara, K., Tajika, A., Imai, H., Takeshima, N., Hayasaka, Y., & Furukawa, T. A. (2015). Protocol registration and selective outcome reporting in recent psychiatry trials: New antidepressants and cognitive behavioural therapies. *Acta Psychiatrica Scandinavica*, 132, 489–498.
- Silove, D., Liddell, B., Rees, S., Chey, T., Nickerson, A., Tam, N., et al. (2014). Effects of recurrent violence on post-traumatic stress disorder and severe distress in conflict-affected timor-leste: A 6-year longitudinal study. *Lancet Glob Health*, 2, e293–300.
- Skapinakis, P., Caldwell, D. M., Hollingworth, W., Bryden, P., Fineberg, N. A., Salkovskis, P., et al. (2016). Pharmacological and psychotherapeutic interventions for management of obsessive-compulsive disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 3, 730–739.
- Slade, E., Keeney, E., Mavranzeouli, I., Dias, S., Fou, L., Stockton, S., et al. (2018). Treatments for bulimia nervosa: A network meta-analysis. *Psychological Medicine*, 48,

- 2629–2636.
- Smit, Y., Huibers, M. J., Ioannidis, J. P. A., van Dyck, R., van Tilburg, W., & Arntz, A. (2012). The effectiveness of long-term psychoanalytic psychotherapy—a meta-analysis of randomized controlled trials. *Clinical Psychology Review, 32*, 81–92.
- Steinert, C., Stadter, K., Stark, R., & Leichsenring, F. (2017). The effects of waiting for treatment: A meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. *Clinical Psychology & Psychotherapy, 24*, 649–660.
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S. L., Krueger, R. F., Miller, J. D., et al. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science, 12*, 742–756.
- Tannenbaum, S., Ross, J. S., Krumholz, H. M., Desai, N. R., Ritchie, J. D., Lehman, R., et al. (2018). Early experiences with journal data sharing policies: A survey of published clinical trial investigators. *Annals of Internal Medicine, 169*(8), 586–588. <https://doi.org/10.7326/M18-0723>.
- Thielking, M. (2019). To improve mental health treatments, scientists try to dissect the pieces that make them work. *STAT*. <https://www.statnews.com/2019/02/01/back-translation-mental-health-treatments/>, Accessed date: 24 July 2019.
- A-Tjak, J., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A., et al. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics, 84*, 30–36.
- Turner, E. H., Matthews, A. M., Linardatos, E., Tell, R. A., & Rosenthal, R. (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine, 358*, 252–260.
- Turner, D. T., van der Gaag, M., Karyotaki, E., & Cuijpers, P. (2014). Psychological interventions for psychosis: A meta-analysis of comparative outcome studies. *American Journal of Psychiatry, 171*, 523–538.
- Van Bockstaele, B., Verschuere, B., Tibboel, H., De Houwer, J., Crombez, G., & Koster, E. H. (2014). A review of current evidence for the causal impact of attentional bias on fear and anxiety. *Psychological Bulletin, 140*, 682–721.
- Walker, M. P., Brakefield, T., Hobson, J. A., & Stickgold, R. (2003). Dissociable stages of human memory consolidation and reconsolidation. *Nature, 425*, 616–620.
- Wallach, J. D., Boyack, K. W., & Ioannidis, J. P. A. (2018). Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. *PLoS Biology, 16*, e2006930.
- Wilson, G. T., Wilfley, D. E., Agras, W. S., & Bryson, S. W. (2011). Allegiance bias and therapist effects: Results of a randomized controlled trial of binge eating disorder. *Clinical Psychologist, 18*, 119–125.
- Yoder, W. R., Karyotaki, E., Cristea, I.-A., van Duin, D., & Cuijpers, P. (2019). Researcher allegiance in research on psychosocial interventions: meta-research study protocol and pilot study. *BMJ Open, 9* bmjopen-2018-024622.