**DPSPs Internationale seminar i København**
**Fredag d. 7. oktober 2016**

Bobby Zachariae
Professor, dr.med., cand.psych.
Aarhus Universitetshospital og
Aarhus Universitet

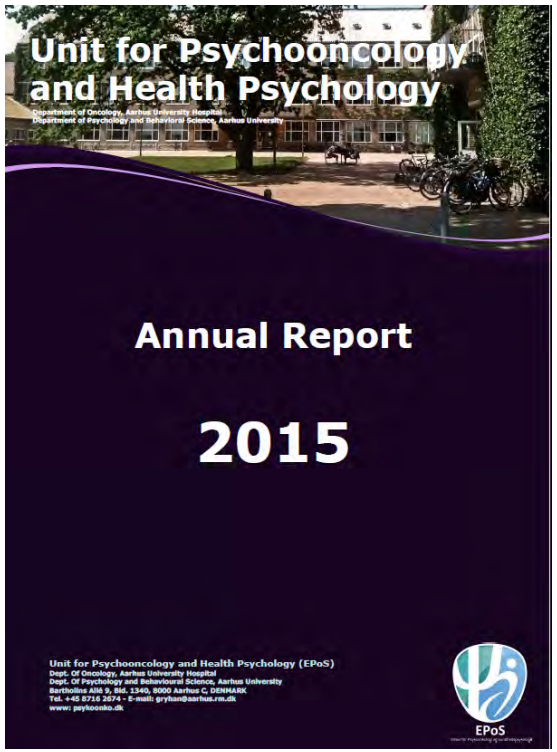
# Forskningsdesigns og evidens

# Aims of the presentation

- Evaluating the available evidence – challenges and available solutions
  - Problems in every-day thinking
  - Aims of scientific method
  - Types of evidence
  - Threats to validity
  - When one study is not enough
  - Aspects to consider when evaluating "effectiveness"

# Unit for Psychooncology and Health Psychology - EPoS

- Psychooncology Research Unit established in 2000 at the Dept. Of Oncology, AUH based on a grant from the Danish Cancer Society
- EPoS established in 2011 in collaboration between AUH, Dept. Of Oncology, BSS, AU, and Dept of Psychology and Behavioural Science
- Current staff: 17 (1 professor, 2 assoc. prof. 1 assist prof, 1 senior researcher, 4 post-docs, 7 PhD's, 1 adm.) + 8-10 research assistants.



# My background

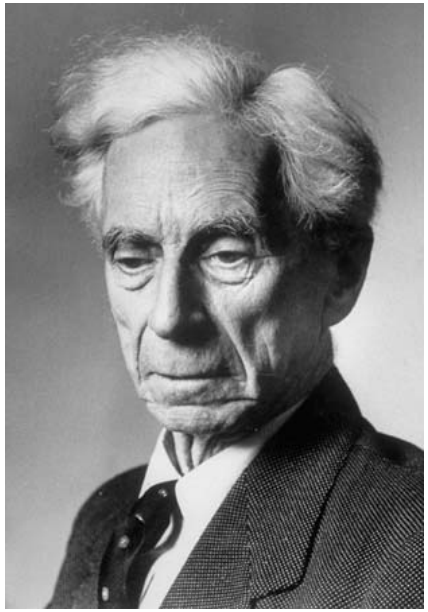
- Research areas
  - Psychoneuroimmunology
  - Pain research
  - Psycho-social cancer research
  - Patient-health professional interactions
  - Health psychology
  - Psychosocial interventions
  - Hypnosis, guided imagery, mindfulness-based intervention, expressive writing
  - Internet-delivered interventions
  - Efficacy of complementary and alternative treatments

# My background

- Research methodologies
  - Experimental studies (psychophysiology)
  - Randomized Controlled Trials (RCT)
  - Pragmatic trials
  - Mixed (qualitative and quantitative) methods
  - Cross-sectional and cohort studies
  - Psychometrics
  - Systematic reviews and meta-analysis

- **Intellectual advise to future generations:**

- *"Ask yourself only what are the facts and what is the truth that the facts bear out. Never let yourself be diverted either by what you wish to believe, or by what you think would have beneficent social effects if it were believed. But look only, and solely, at what are the facts."*



Bertrand Russell, 1872-1970

# Why scientific method?

*Personal experience*

*vs*

*Scientific evidence*

*”Man prefers to believe what he prefers to be true”*

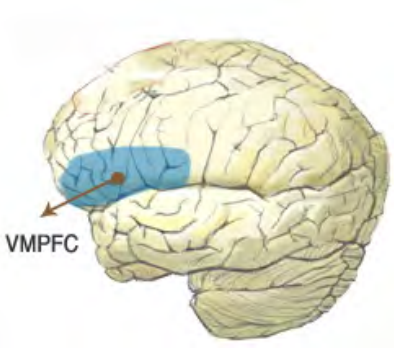
Francis Bacon, 1561-1626

# Automatic vs manual processing
# "Fast" vs "slow" thinking

**System 1**
Fast
Automatic
Emotional
Frequent
Stereotypic
Subconscious

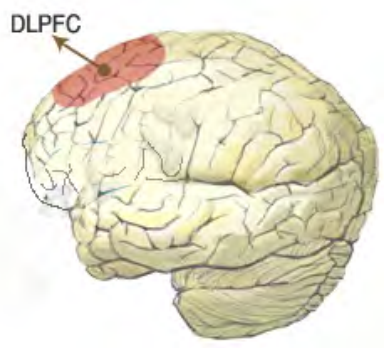
**Evolution has equipped humans with two types of thinking**

**System 2**
Slow
Effortful
Logical
Infrequent
Calculating
Conscious




Automatic mode

Ventromedial prefrontal cortex




Manual mode

Dorsolateral prefrontal cortex

Shiv et al. 2005; Kahneman, 2011

# Problems in Everyday- (fast) Thinking

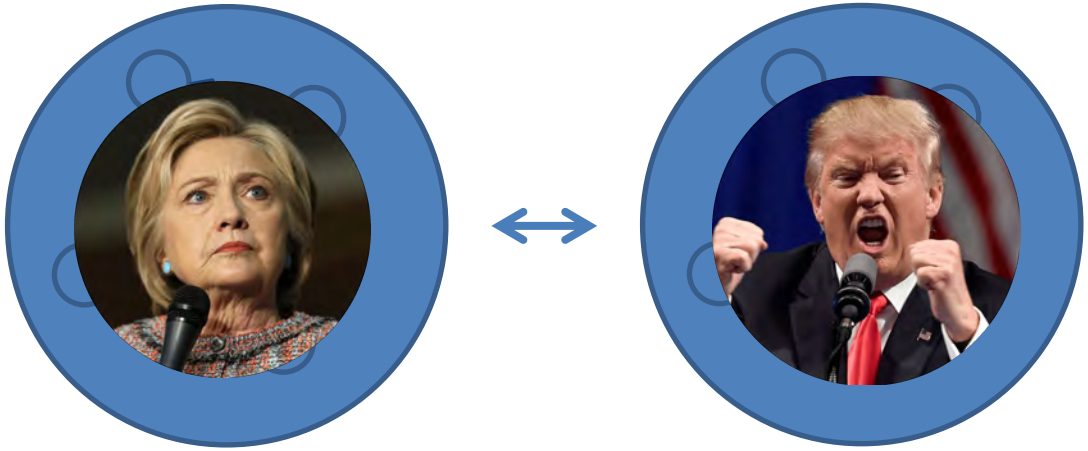
- The practical value of everyday (”fast”) thinking is obtained at the cost of bias and precision, e.g.,:
  - We tend to see patterns, even where there are none
  - We see causal relationships, even where there are none
  - We tend to focus on and remember positive evidence
  - We tend to overestimate evidence confirming our position
  - Our judgments are influenced by the judgments of our surroundings (conformity)
  - We tend to believe that positive and negative traits , respectively, are associated (clustering)
  - We tend to overestimate the probability of dramatic events

# Problems in fast thinking

- The practical value of everyday (”fast”) thinking is obtained at the cost of bias and precision, e.g.,
  - We have a tendency to see patterns, also when there are no patterns:
    - E.g. in completely random sequences: 12221222122112112211.
    - E.g. When rolling dice: If we have not obtained a ”six” in many rolls, we tend to believe that the probability increases (although the chance remains 1/6 even after 100 rolls)
  - We tend to perceive causal relationships, also when there are none:
    - E.g., we tend to imagine causality between X and Y, if Y takes place after X (a necessary but not sufficient condition for causality)

# The echo chamber problem

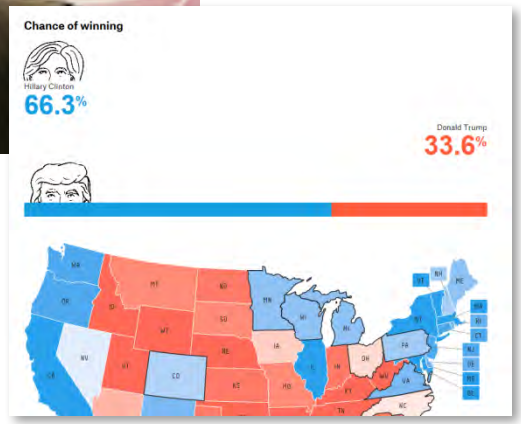
- Closed ideology echo chamber
- Applies to politics as well as other domains



The solution is a no-brainer – but difficult to practice

*”We love to predict things – and we aren’t very good at it”*


Nate Silver (1978 -)




Sept.30

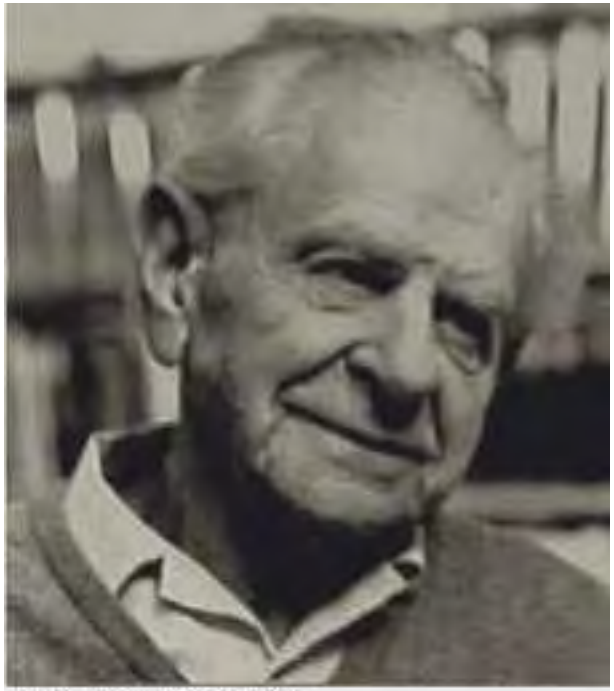
# The domains of science

- Meta-physics: Philosophy, epistemology: deduction and reasoning
- Theories of science: Meta-theories about method
- Theoretical science: Collecting, condensing, discussing, and interpreting existing theoretical and empirical research results
- Empirical science: Measuring phenomena and testing hypotheses
    - **Observing and describing**
    - **Predicting**
    - **Determining causes**
    - **Explaining**

# The aim of scientific method

- **General aim**: To generate measurable and testable data, gradually adding to the accumulation of human knowledge
  - To produce reliable knowledge
  - To produce valid knowledge
  - About causal relationships
  - By addressing possible sources of error

# The falsification principle



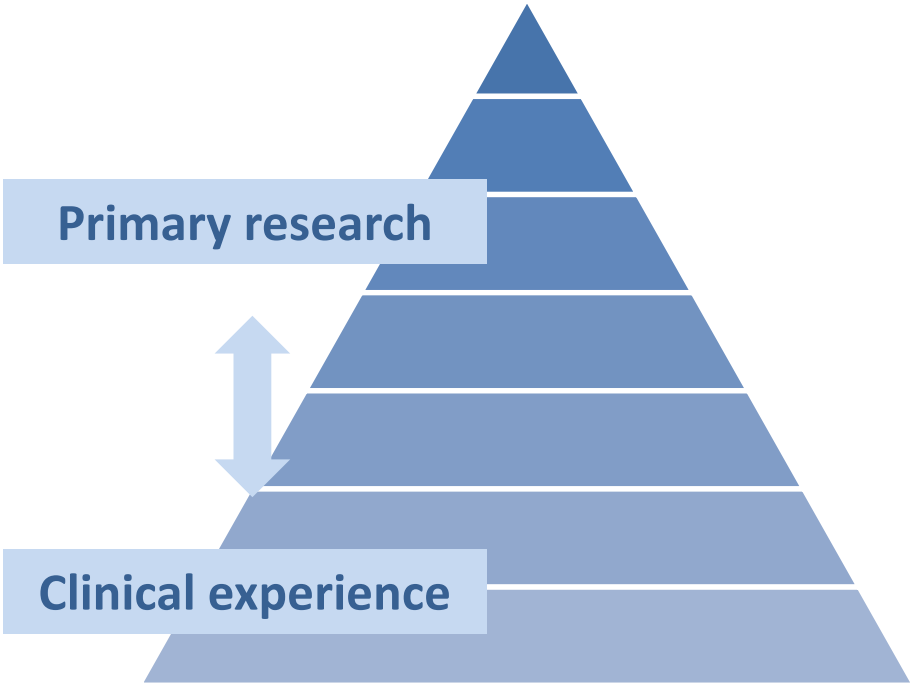
Sir Karl Popper (1902-1994)


**Karl Popper: A Scientific Hypothesis Must Be "Falsifiable".**

- *We support a hypothesis by falsifying the null-hypothesis*
- *A general approach: We do not "prove" hypotheses – but maximize our attempts to falsify statements about observations, associations, causality, and mechanisms*

# Types of evidence

# Evidence hierarchy



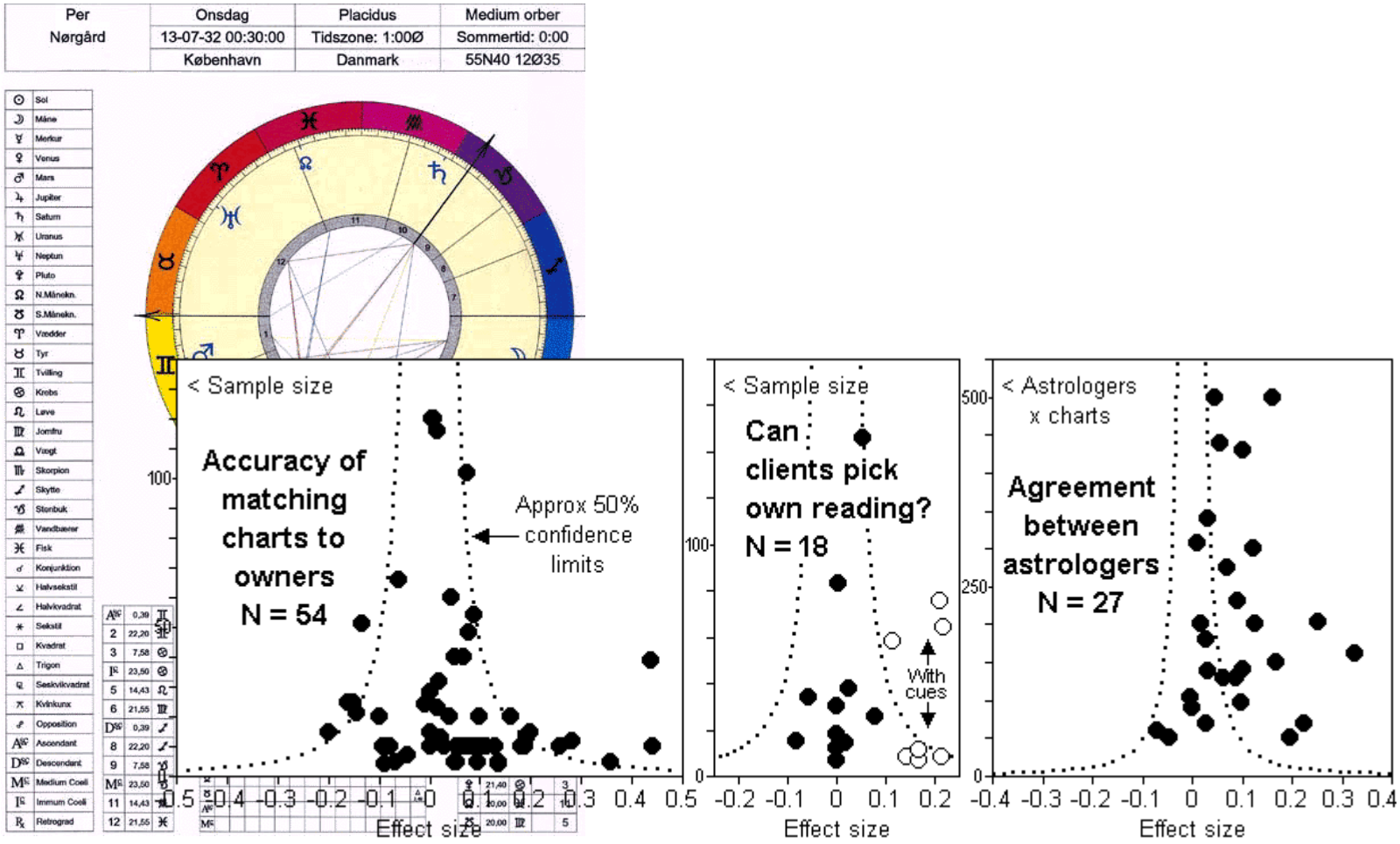

- Experimental design (RCT)
- Cohort study
- Case-control study
- Cross sectional study
- Case reports
- Expert opinion
- Anecdote

Challenge: the model favors internal validity

# Types of validity

- Validity
  - Internal validity (causality, excluding alternative explanations, sources of error)
  - External validity (generalizability)
  - Ecological validity (pragmatic validity)
- The three types of validity supplement each other
- Are difficult to obtain with one single method
- Internal validity a prerequisite for external and ecological validity
- Reliability a prerequisite for validity – but not the reverse

# Mistaking reliability for validity



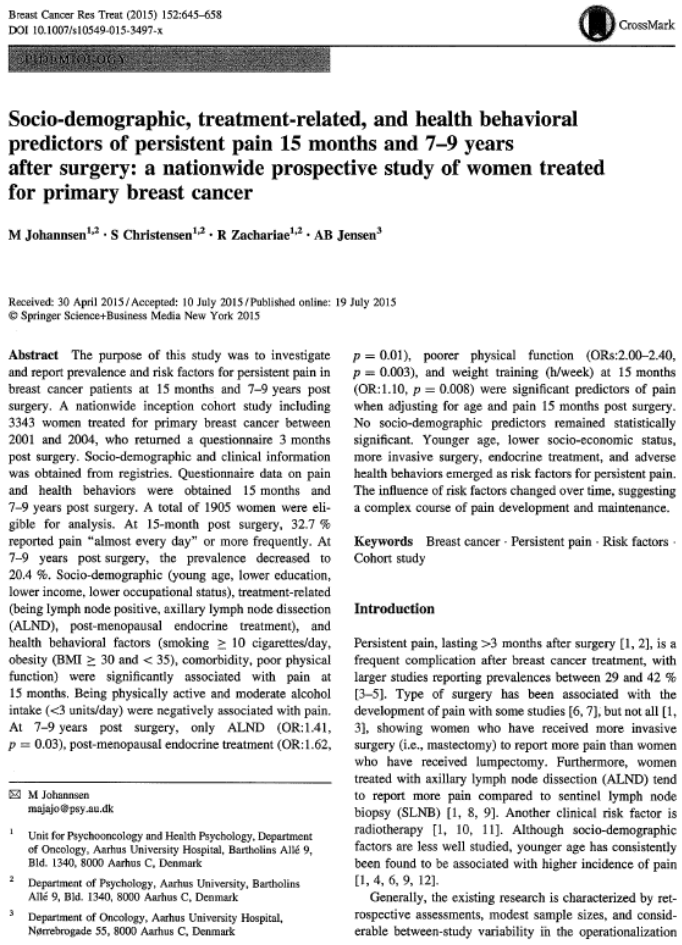
# Research questions

- Internal validity
  - Does it work (statistical significance, superiority)?
  - How well does it work (practical significance)?
  - Does it work as well as something else (non-inferiority)?
  - How does it work (mechanisms, specific, non-specific)?
- External validity
  - For whom does it work?
  - For whom does it not work?
- Ecological validity
  - Does it work in the clinical practical context?
  - Clinician and patient adherence

# Example of design *maximizing internal validity:* Randomized controlled trial

# MBCT for persistent pain in women treated for breast cancer

- 16-20% of women treated for breast cancer experience pain after 5-9 years
- Limited pharmacological treatment efficacy
- Pain is a multidimensional phenomenon consisting of sensory, cognitive, and affective factors
- Mindfulness-based therapy teaches ways of relating to bodily sensations and emotional discomfort with higher degree of acceptance and openness
- Mindfulness-Based Cognitive Therapy may be effective for cancer-related pain



Breast Cancer Res Treat (2015) 152:645–658
DOI 10.1007/s10549-015-3497-x

EPIDEMIOLOGY

**Socio-demographic, treatment-related, and health behavioral predictors of persistent pain 15 months and 7–9 years after surgery: a nationwide prospective study of women treated for primary breast cancer**

M Johannsen[1,2] · S Christensen[1,2] · R Zachariae[1,2] · AB Jensen[3]

Received: 30 April 2015 / Accepted: 10 July 2015 / Published online: 19 July 2015


**Abstract** The purpose of this study was to investigate and report prevalence and risk factors for persistent pain in breast cancer patients at 15 months and 7–9 years post surgery. A nationwide inception cohort study including 3343 women treated for primary breast cancer between 2001 and 2004, who returned a questionnaire 3 months post surgery. Socio-demographic and clinical information was obtained from registries. Questionnaire data on pain and health behaviors were obtained 15 months and 7–9 years post surgery. A total of 1905 women were eligible for analysis. At 15-month post surgery, 32.7 % reported pain "almost every day" or more frequently. At 7–9 years post surgery, the prevalence decreased to 20.4 %. Socio-demographic (young age, lower education, lower income, lower occupational status), treatment-related (being lymph node positive, axillary lymph node dissection (ALND), post-menopausal endocrine treatment), and health behavioral factors (smoking ≥ 10 cigarettes/day, obesity (BMI ≥ 30 and < 35), comorbidity, poor physical function) were significantly associated with pain at 15 months. Being physically active and moderate alcohol intake (<3 units/day) were negatively associated with pain. At 7–9 years post surgery, only ALND (OR:1.41, $p = 0.03$), post-menopausal endocrine treatment (OR:1.62, $p = 0.01$), poorer physical function (ORs:2.00–2.40, $p = 0.003$), and weight training (h/week) at 15 months (OR:1.10, $p = 0.008$) were significant predictors of pain when adjusting for age and pain 15 months post surgery. No socio-demographic predictors remained statistically significant. Younger age, lower socio-economic status, more invasive surgery, endocrine treatment, and adverse health behaviors emerged as risk factors for persistent pain. The influence of risk factors changed over time, suggesting a complex course of pain development and maintenance.

**Keywords** Breast cancer · Persistent pain · Risk factors · Cohort study

**Introduction**

Persistent pain, lasting >3 months after surgery [1, 2], is a frequent complication after breast cancer treatment, with larger studies reporting prevalences between 29 and 42 % [3–5]. Type of surgery has been associated with the development of pain with some studies [6, 7], but not all [1, 3], showing women who have received more invasive surgery (i.e., mastectomy) to report more pain than women who have received lumpectomy. Furthermore, women treated with axillary lymph node dissection (ALND) tend to report more pain compared to sentinel lymph node biopsy (SLNB) [1, 8, 9]. Another clinical risk factor is radiotherapy [1, 10, 11]. Although socio-demographic factors are less well studied, younger age has consistently been found to be associated with higher incidence of pain [1, 4, 6, 9, 12].

Generally, the existing research is characterized by retrospective assessments, modest sample sizes, and considerable between-study variability in the operationalization

✉ M Johannsen
majajo@psy.au.dk

[1] Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital, Bartholins Allé 9, Bld. 1340, 8000 Aarhus C, Denmark

[2] Department of Psychology, Aarhus University, Bartholins Allé 9, Bld. 1340, 8000 Aarhus C, Denmark

[3] Department of Oncology, Aarhus University Hospital, Nørrebrogade 55, 8000 Aarhus C, Denmark

# MBCT for persistent pain in women treated for breast cancer

JOURNAL OF CLINICAL ONCOLOGY | ORIGINAL REPORT

## Efficacy of Mindfulness-Based Cognitive Therapy on Late Post-Treatment Pain in Women Treated for Primary Breast Cancer: A Randomized Controlled Trial

*Maja Johannsen, Maja O'Connor, Mia Skytte O'Toole, Anders Bonde Jensen, Inger Højris, and Robert Zachariae*

Maja Johannsen, Maja O'Connor, Mia Skytte O'Toole, Anders Bonde Jensen, Inger Højris, and Robert Zachariae, Aarhus University Hospital; Anders Bonde Jensen and Inger Højris, Aarhus University, Aarhus, Denmark.

Published online ahead of print at www.jco.org on June 20, 2016.

Supported by The Danish Cancer Society; Aase & Einar Danielsens Fond, Einar Willumsens Mindelegat, and Radiumstationens Forskningsfond.

Authors' disclosures of potential conflicts of interest are found in the article online at www.jco.org. Author contributions are found at the end of this article.

Clinical trial information: NCT01674881.

Corresponding author: Maja Johannsen, MSc, Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital and Department of Psychology, Aarhus University, Bartholins Allé 9, Bld 1340, DK-8000 Aarhus C, Denmark; e-mail: majajo@psy.au.dk.



0732-183X/16/3499-1/$20.00

DOI: 10.1200/JCO.2015.65.0770

**ABSTRACT**

**Purpose**
To assess the efficacy of mindfulness-based cognitive therapy (MBCT) for late post-treatment pain in women treated for primary breast cancer.

**Methods**
A randomized wait list–controlled trial was conducted with 129 women treated for breast cancer reporting post-treatment pain (score ≥ 3 on pain intensity or pain burden assessed with 10-point numeric rating scales). Participants were randomly assigned to a manualized 8-week MBCT program or a wait-list control group. Pain was the primary outcome and was assessed with the Short Form McGill Pain Questionnaire 2 (SF-MPQ-2), the Present Pain Intensity subscale (the McGill Pain Questionnaire), and perceived pain intensity and pain burden (numeric rating scales). Secondary outcomes were quality of life (World Health Organization-5 Well-Being Index), psychological distress (the Hospital Depression and Anxiety Scale), and self-reported use of pain medication. All outcome measures were assessed at baseline, postintervention, and 3-month and 6-month follow-up. Treatment effects were evaluated with mixed linear models.

**Results**
Statistically significant time × group interactions were found for pain intensity ($d = 0.61$; $P = .002$), the Present Pain Intensity subscale ($d = 0.26$; $P = .026$), the SF-MPQ-2 neuropathic pain subscale ($d = 0.24$; $P = .036$), and SF-MPQ-2 total scores ($d = 0.23$; $P = .036$). Only pain intensity remained statistically significant after correction for multiple comparisons. Statistically significant effects were also observed for quality of life ($d = 0.42$; $P = .028$) and nonprescription pain medication use ($d = 0.40$; $P = .038$). None of the remaining outcomes reached statistical significance.

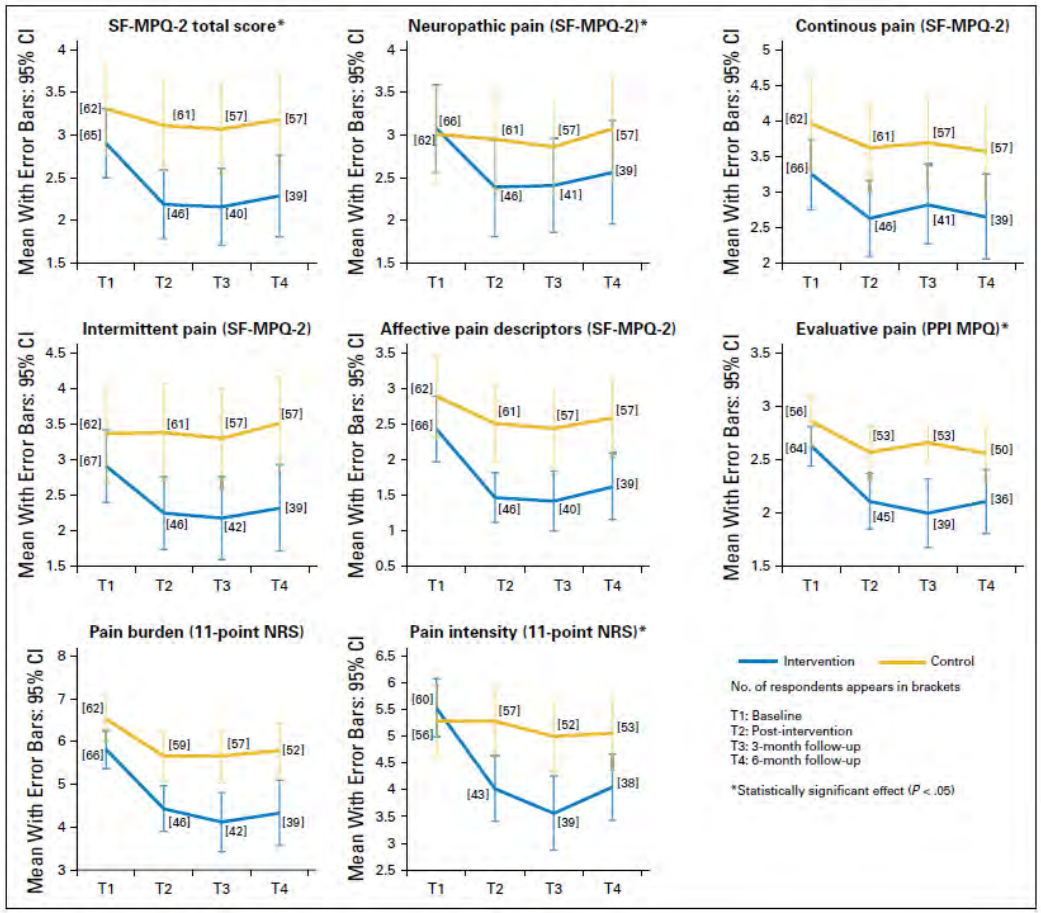
**Conclusion**
MBCT showed a statistically significant, robust, and durable effect on pain intensity, indicating that MBCT may be an efficacious pain rehabilitation strategy for women treated for breast cancer. In addition, the effect on neuropathic pain, a pain type reported by women treated for breast cancer, further suggests the potential of MBCT but should be considered preliminary.

*J Clin Oncol 34. © 2016 by American Society of Clinical Oncology*

**INTRODUCTION**

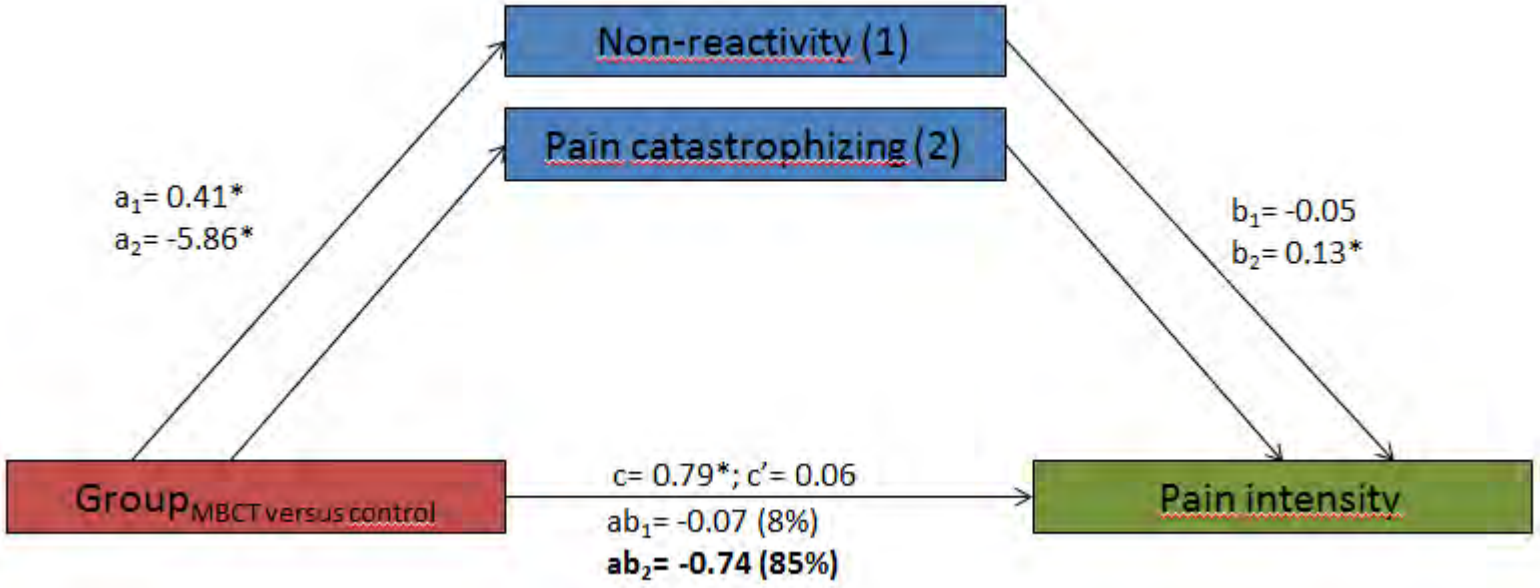
Breast cancer (BC) is the most common cancer type among women, with > 1 million new cases worldwide each year.[1] Although survival rates are increasing,[2] patients with BC report high levels of physical and psychological symptoms after treatment,[3,4] underscoring the need for effective rehabilitation programs. Persistent post-treatment pain is of particular concern because of its high prevalence, with 16% to 20% of women treated for BC experiencing moderate to severe pain 5 to 9 years after surgery.[5,6] Although women treated for BC rate pain among the 10 most important quality of life (QoL) –related issues,[7] pain remains undertreated.[8,9]

Although BC-related pain is often associated with nervous system damage (ie, neuropathic pain),[10] it is generally accepted that pain is multifaceted, consisting of sensory, cognitive, and affective dimensions influencing the patients' pain experience and their reactions to this experience.[9] Targeting cognitive and affective dimensions of pain with psychosocial interventions may, as supported by a recent meta-analysis,[11] help patients with BC cope better with persisting post-treatment pain. One type of intervention,



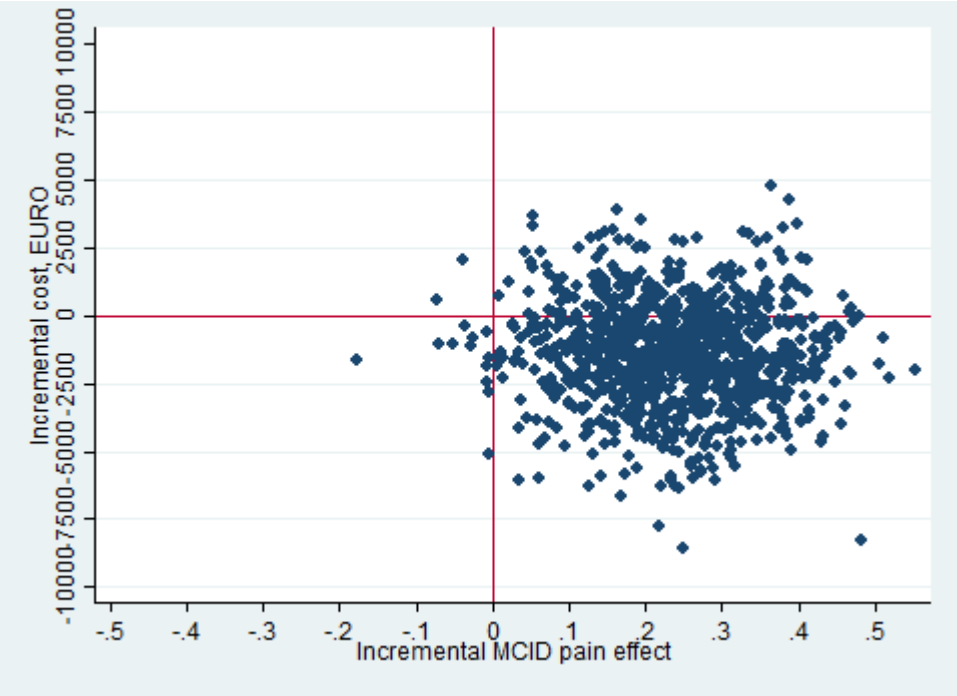
# MBCT for persistent pain in women treated for breast cancer

- *Mediators* of the effect of MBCT on pain intensity:
  - Mindfulness non-reactivity facet; Pain catastrophizing




**Johannsen, O'Toole, O'Connor, Jensen & Zachariae (under review)**

# MBCT for persistent pain in women treated for breast cancer




Cost-effectiveness of MBCT for persistent pain in women treated for breast cancer

**Johannsen, Sørensen, O'Connor, O'Toole, Zachariae (in preparation)**

# Example of design
# Pragmatic trial maximizing both *internal, external, and ecological* validity

# Evaluating complementary and alternative treatments

- Lab experiments and RCTs of energy healing yield negative results

  Zachariae et al. 2005

- Complementary and alternative medicines and treatments (CAMs) are associated with increased symptoms of depression in breast cancer patients

  Pedersen et al. 2013

- Active acupuncture is not more effective than placebo acupuncture in a double blinded RCT

  Vase et al. 2013


British Journal of Cancer (2005) 93, 538–543

www.bjcancer.com

The effect of spiritual healing on *in vitro* tumour cell proliferation and viability – an experimental study

R Zachariae*

Breast Cancer Res Treat (2013) 141:277–285

EPIDEMIOLOGY

**Use of complementary and alternative medicine (CAM) and changes in depressive symptoms from 3 to 15 months after surgery for primary breast cancer: results from a nationwide cohort study**

NIH Public Access

Author Manuscript

Published in final edited form as:

*Pain*. 2013 September ; 154(9): 1659–1667. doi:10.1016/j.pain.2013.05.008.

**Specifying the non-specific components of acupuncture analgesia**

Lene Vase[a,b,*], Sara Baram[c,d], Nobuari Takakura[e], Hiroyoshi Yajima[e], Miho Takayama[e], Ted J. Kaptchuk[f], Søren Schou[g], Troels Staehelin Jensen[b], Robert Zachariae[a,g], and Peter Svensson[c]

[a]Department of Psychology, Aarhus University, Aarhus, Denmark

[b]Danish Pain Research Center, Aarhus University Hospital, Aarhus, Denmark

[c]Section of Clinical Oral Physiology, Department of Dentistry, Aarhus University, Aarhus, Denmark

[d]Section of Oral and Maxillofacial Surgery and Oral Pathology, Department of Dentistry, Aarhus University, Aarhus, Denmark

[e]Department of Acupuncture and Moxibustion, Tokyo Ariake University of Medical and Health Sciences, Tokyo, Japan

[f]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA

[g]Unit for Psychooncology and Health Psychology, Aarhus University Hospital, Aarhus, Denmark

**Abstract**

It is well known that acupuncture has pain-relieving effects, but the contribution of specific and especially non-specific factors to acupuncture analgesia is less clear. One hundred and one patients who developed pain ≥ 3 on a visual analog scale (VAS, 0-10) following third molar surgery were randomized to receive active acupuncture, placebo acupuncture, or no treatment for 30 min with acupuncture needles with potential for double-blinding. Patients' perception of the treatment (active or placebo), and expected pain levels (VAS) were assessed prior to and halfway through the treatment. Looking at actual treatment allocation, there was no specific effect of active acupuncture ($P = 0.240$), but a large and significant non-specific effect of placebo acupuncture ($P < 0.001$), which increased over time. Interestingly, however, looking at perceived treatment allocation, there was a significant effect of acupuncture ($P < 0.001$) indicating that patients who believed they received active acupuncture had significantly lower pain levels than those who believed they received placebo acupuncture. Expected pain levels accounted for significant and progressively larger amounts of the variance in pain ratings following both active and placebo acupuncture (up to 69.8%). This is the first study to show that under optimized blinding conditions non-specific factors such as patients' perception of and expectations toward treatment are central to the efficacy of acupuncture analgesia and that these factors may contribute to self-reinforcing effects in acupuncture treatment To obtain an effect of acupuncture in clinical practice it may, therefore, be important to incorporate and optimize these factors.

# Effectiveness of energy healing on quality of life in colorectal cancer patients

- CAM practitioners criticize validity of traditional research methodologies, e.g., RCTs
- Criticisms include:
  - May not be generalized to the general population
  - People may have strong treatment preferences
  - Standardized outcome measures may not cover patients' individual concerns
  - Patients may prefer some practitioners to others
  - Standardization of treatment context may cancel out effects

# Effectiveness of energy healing on quality of life in colorectal cancer patients

Complementary Therapies in Medicine (2014) 22, 463–472

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevierhealth.com/journals/ctim

Effectiveness of energy healing on Quality of Life: A pragmatic intervention trial in colorectal cancer patients☆

Christina Gundgaard Pedersen[a,*], Helle Johannessen[b], Jacob V.B. Hjelmborg[c], Robert Zachariae[a]

[a] Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital and Department of Psychology and Behavioral Science, Aarhus University, Bartholins Alle 9, 8000 Aarhus C, Denmark
[b] Health, Man and Society, Institute of Public Health, University of Southern Denmark, J.B. Winsløws Vej 9B, 5000 Odense C, Denmark
[c] Department of Biostatistics, Institute of Public Health, University of Southern Denmark, J.B. Winsløws Vej 9B, 5000 Odense C, Denmark

Available online 2 May 2014

KEYWORDS
Complementary and alternative medicine (CAM);
Cancer;
Healing;
Quality of Life (QoL);
Sleep quality;
Depressive symptoms;
Mood

Summary
*Purpose:* Our aim was to explore the effectiveness of energy healing, a commonly used complementary and alternative therapy, on well-being in cancer patients while assessing the possible influence on the results of participating in a randomized controlled trial.
*Methods:* 247 patients treated for colorectal cancer (response rate: 31.5%) were either (a) randomized to healing (RH) or control (RC) or (b) had self selected the healing (SH) or control condition (SC), and completed questionnaires assessing well-being (QoL, depressive symptoms, mood, and sleep quality), attitude toward complementary and alternative medicine (CAM), and faith/spirituality at baseline, 1 week, and 2 months post-intervention. They also indicated, at baseline, whether they considered QoL, depressive symptoms, mood, and sleep quality as important outcomes to them.

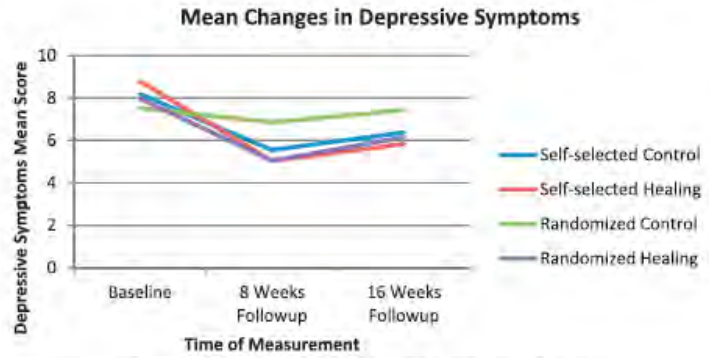
☆ Financial support was provided by The Danish Council for Strategic Research, Ministry of Science, Innovation and Higher Education (Grant no. 09-065176); Department of Psychology and Behavioral Sciences, University of Aarhus; and the Institute of Clinical Research and Institute of Public Health, University of Southern Denmark. The funding sources had no involvement in the study design, collection, analysis, and interpretation of data, or writing of the paper.
* Corresponding author. Tel.: +45 87 16 58 79.
*E-mail address:* christina@psy.au.dk (C.G. Pedersen).

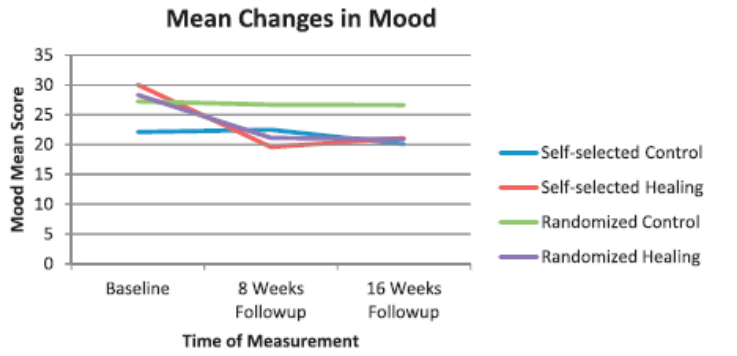
http://dx.doi.org/10.1016/j.ctim.2014.04.003
0965-2299/

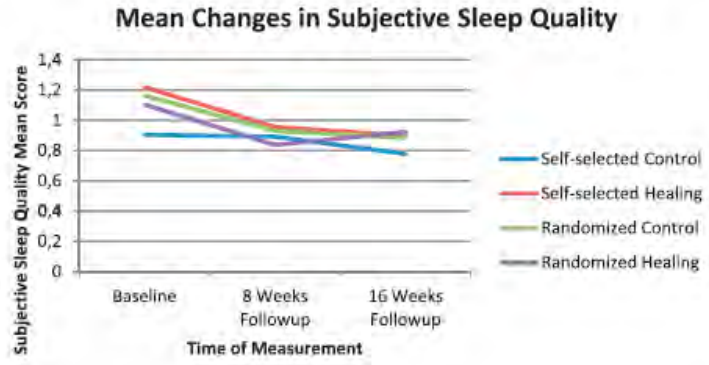
- Study designed to maximize internal, external, and ecological validity
- Colorectal cancer patients randomized to:
  - A Randomization
    - Healing or control
  - B Self-selection
    - Healing or control
- Patients
  - Selected their healer from a list
  - Treatment took place in healer's clinic
  - Completed standardized QoL measures
  - Prioritized preferred outcome
  - Completed measure of attitude towards CAM

# Effectiveness of energy healing on quality of life in colorectal cancer patients

- No overall effects on any outcomes
- Small effect on QoL in subgroup: Patients in self-selected healing group who had rated QoL as important, and who had a positive attitude towards CAM

# Example of design
# Pragmatic evaluation of daily practice


Robert Zachariae
Professor, MSc. DMSci
bzach@aarhus.rm.dk

# Pragmatic evaluation

- Fagligt Selskab for Psykologer i Palliation og Onkologi
- 11 psychologists treating 92 patients or caregivers
- Psychologist questionnaires: psychotherapeutic models and tools used
- Patient questionnaires pre- and post consulation:
  - MYCaW (Measure Yourself Concerns and Well-being): Primary and secondary concern and general well-being
  - Working Alliance Inventory
- N of 1 statistics: Reliable Change Index (RCI)
  - Determines whether a change is beyond a statistical error

**Pedersen et al. (unpublished)**

# Pragmatic evaluation

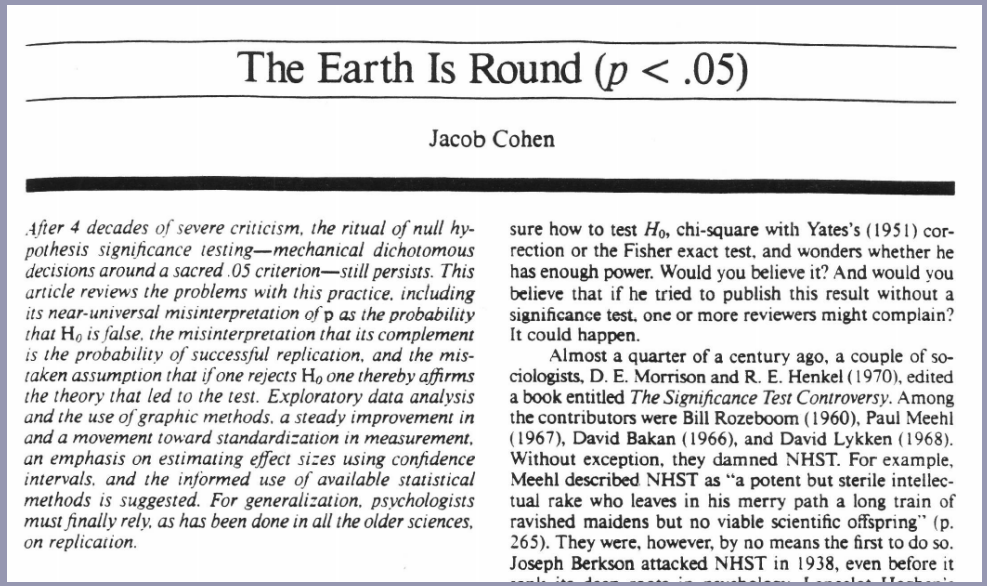
| Selvvalgt problemområde og alment velbefindende | Sign. (RCI) Forbedring | Ingen ændring | Sign. (RCI) Forværring | Data mangler |
|---|---|---|---|---|
| | N (%) | N (%) | N (%) | N (%) |
| Primært problemområde | 22 (23,9) | 64 (69,6) | 0 (0,0) | 6 (6,5) |
| Sekundært problemområde | 20 (21,7) | 43 (46,7) | 0 (0,0) | 29 (31,5) |
| Alment velbefindende | 28 (30,4) | 60 (65,2) | 1 (1,1) | 3 (3,3) |

**Statistically significant predictors of sign. improvement:**

- Positive expectancies: The session will improve my understanding of my reactions and emotions
- Perceived working alliance
- Higher educational level

**Pedersen et al. (unpublished)**

# Interpreting results



## The Earth Is Round ($p < .05$)

Jacob Cohen

*After 4 decades of severe criticism, the ritual of null hypothesis significance testing—mechanical dichotomous decisions around a sacred .05 criterion—still persists. This article reviews the problems with this practice, including its near-universal misinterpretation of* p *as the probability that* $H_0$ *is false, the misinterpretation that its complement is the probability of successful replication, and the mistaken assumption that if one rejects* $H_0$ *one thereby affirms the theory that led to the test. Exploratory data analysis and the use of graphic methods, a steady improvement in and a movement toward standardization in measurement, an emphasis on estimating effect sizes using confidence intervals, and the informed use of available statistical methods is suggested. For generalization, psychologists must finally rely, as has been done in all the older sciences, on replication.*

sure how to test $H_0$, chi-square with Yates's (1951) correction or the Fisher exact test, and wonders whether he has enough power. Would you believe it? And would you believe that if he tried to publish this result without a significance test, one or more reviewers might complain? It could happen.

Almost a quarter of a century ago, a couple of sociologists, D. E. Morrison and R. E. Henkel (1970), edited a book entitled *The Significance Test Controversy*. Among the contributors were Bill Rozeboom (1960), Paul Meehl (1967), David Bakan (1966), and David Lykken (1968). Without exception, they damned NHST. For example, Meehl described NHST as "a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no viable scientific offspring" (p. 265). They were, however, by no means the first to do so. Joseph Berkson attacked NHST in 1938, even before it

# What is an effect size?

A standardized effect, e.g., standardized mean difference, enabling comparisons across measures and studies

**Cohen's d = (Mean 1 – Mean 2)/SD (pooled)**

**Which intervention is most effective?**

- Study 1: Mean score and standard deviation of Hamilton Depression Rating Scale (range: 0-49):
- Intervention: 16.5 (13.0)
- Control: 20.5 (14.0)
- Cohen's d = **0.29**

**HDRS MCID (0,5 SD) [1]**

- Study 2: Mean score and standard deviation of Beck's Depression Inventory (range: 0-63)
- Intervention: 17.5 (7.0)
- Control: 21.5 (8.0)
- Cohen's d = **0.53**

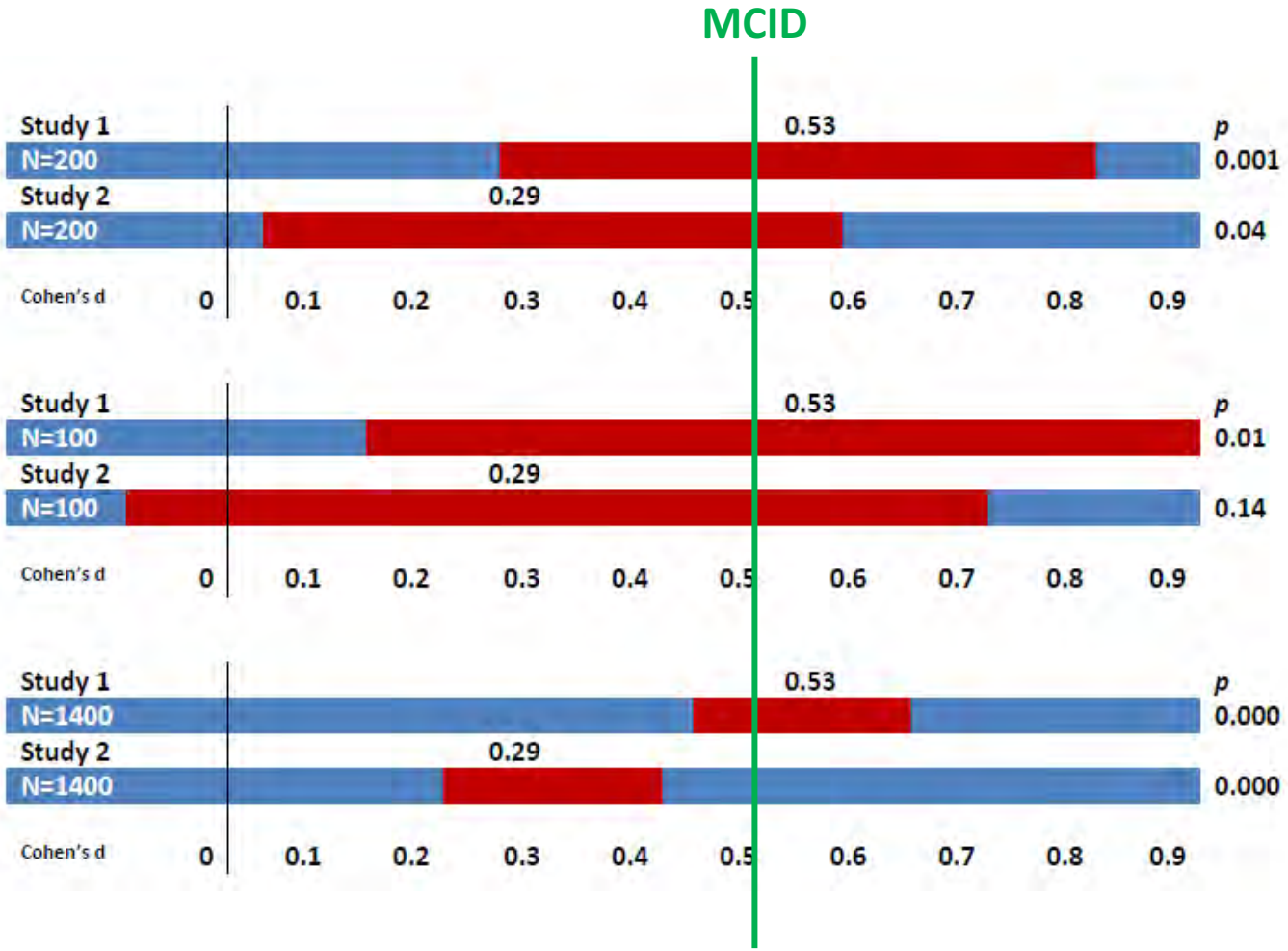
**BDI MCID (17%) [2] = 3.7 = SD: 0.49**

*) To detect the difference in *d* between study 1 and 2 requires a sample of 610 in each group



[1] NICE, 2004; [2] Button et al. 2015

# Significance and precision




**Both effect sizes are statistically significantly different from ”0”, are not different from each other**

**Only the effect size of study 1 is significantly different from ”0”. The two effect sizes are not different from each other**

**Both effect sizes are statistically significantly different from ”0”, and two effect sizes are sign. different from each other ($p$ = 0.003)**

Significance: p-values < 0.05; Precision: 95% Confidence interval

# One study is not enough!

Challenges:
Non-replication
publication bias
“cherry picking”

Unit for Psychooncology and Health Psychology
AARHUS UNIVERSITY
Aarhus University Hospital
Robert Zachariae
Professor, MSc. DMSci
bzach@aarhus.rm.dk

# Interpreting non-replicated results

Essay

## Why Most Published Research Findings Are False

John P. A. Ioannidis

**Summary**

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser preselection of tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance. Simulations show that for most study designs and settings, it is more likely for a research claim to be false than true. Moreover, for many current scientific fields, claimed research findings may often be simply accurate measures of the prevailing bias. In this essay, I discuss the implications of these problems for the conduct and interpretation of research.

Published research findings are sometimes refuted by subsequent evidence, with ensuing confusion and disappointment. Refutation and controversy is seen across the range of research designs, from clinical trials and traditional epidemiological studies [1–3] to the most modern molecular research [4,5]. There is increasing concern that in modern research, false findings may be the majority or even the vast majority of published research

factors that influence this problem and some corollaries thereof.

### Modeling the Framework for False Positive Findings

Several methodologists have pointed out [9–11] that the high rate of nonreplication (lack of confirmation) of research discoveries is a consequence of the convenient, yet ill-founded strategy of claiming conclusive research findings solely on the basis of a single study assessed by formal statistical significance, typically for a *p*-value less than 0.05. Research is not most appropriately represented and summarized by *p*-values, but, unfortunately, there is a widespread notion that medical research articles

**It can be proven that most claimed research findings are false.**

should be interpreted based only on *p*-values. Research findings are defined here as any relationship reaching formal statistical significance, e.g., effective interventions, informative predictors, risk factors, or associations. "Negative" research is also very useful. "Negative" is actually a misnomer, and the misinterpretation is widespread. However, here we will target relationships that investigators claim exist, rather than null findings.

As has been shown previously, the probability that a research finding is indeed true depends on the prior probability of it being true (before doing the study), the statistical power of the study, and the level of statistical significance [10,11]. Consider a 2 × 2 table in which research findings are compared against the gold standard

is characteristic of the field and can vary a lot depending on whether the field targets highly likely relationships or searches for only one or a few true relationships among thousands and millions of hypotheses that may be postulated. Let us also consider, for computational simplicity, circumscribed fields where either there is only one true relationship (among many that can be hypothesized) or the power is similar to find any of the several existing true relationships. The pre-study probability of a relationship being true is $R/(R+1)$. The probability of a study finding a true relationship reflects the power $1 - \beta$ (one minus the Type II error rate). The probability of claiming a relationship when none truly exists reflects the Type I error rate, $\alpha$. Assuming that $c$ relationships are being probed in the field, the expected values of the 2 × 2 table are given in Table 1. After a research finding has been claimed based on achieving formal statistical significance, the post-study probability that it is true is the positive predictive value, PPV. The PPV is also the complementary probability of what Wacholder et al. have called the false positive report probability [10]. According to the 2 × 2 table, one gets $PPV = (1 - \beta)R/(R - \beta R + \alpha)$. A research finding is thus

**Citation:** Ioannidis JPA (2005) Why most published research findings are false. PLoS Med 2(8):e124.



**Abbreviation:** PPV, positive predictive value

John P. A. Ioannidis is in the Department of Hygiene and Epidemiology, University of Ioannina School of

ORIGINAL CONTRIBUTION

## Contradicted and Initially Stronger Effects in Highly Cited Clinical Research

John P. A. Ioannidis, MD

**Context** Controversy and uncertainty ensue when the results of clinical research on the effectiveness of interventions are subsequently contradicted. Controversies are most prominent when high-impact research is involved.

**Objectives** To understand how frequently highly cited studies are contradicted or find effects that are stronger than in other similar studies and to discern whether specific characteristics are associated with such refutation over time.

**Design** All original clinical research studies published in 3 major general clinical journals or high-impact-factor specialty journals in 1990-2003 and cited more than 1000 times in the literature were examined.

**Main Outcome Measure** The results of highly cited articles were compared against subsequent studies of comparable or larger sample size and similar or better controlled designs. The same analysis was also performed comparatively for matched studies that were not so highly cited.

**Results** Of 49 highly cited original clinical research studies, 45 claimed that the intervention was effective. Of these, 7 (16%) were contradicted by subsequent studies, 7 others (16%) had found effects that were stronger than those of subsequent studies, 20 (44%) were replicated, and 11 (24%) remained largely unchallenged. Five of 6 highly-cited nonrandomized studies had been contradicted or had found stronger effects vs 9 of 39 randomized controlled trials ($P = .008$). Among randomized trials, studies with contradicted or stronger effects were smaller ($P = .009$) than replicated or unchallenged studies although there was no statistically significant difference in their early or overall citation impact. Matched control studies did not have a significantly different share of refuted results than highly cited studies, but they included more studies with "negative" results.

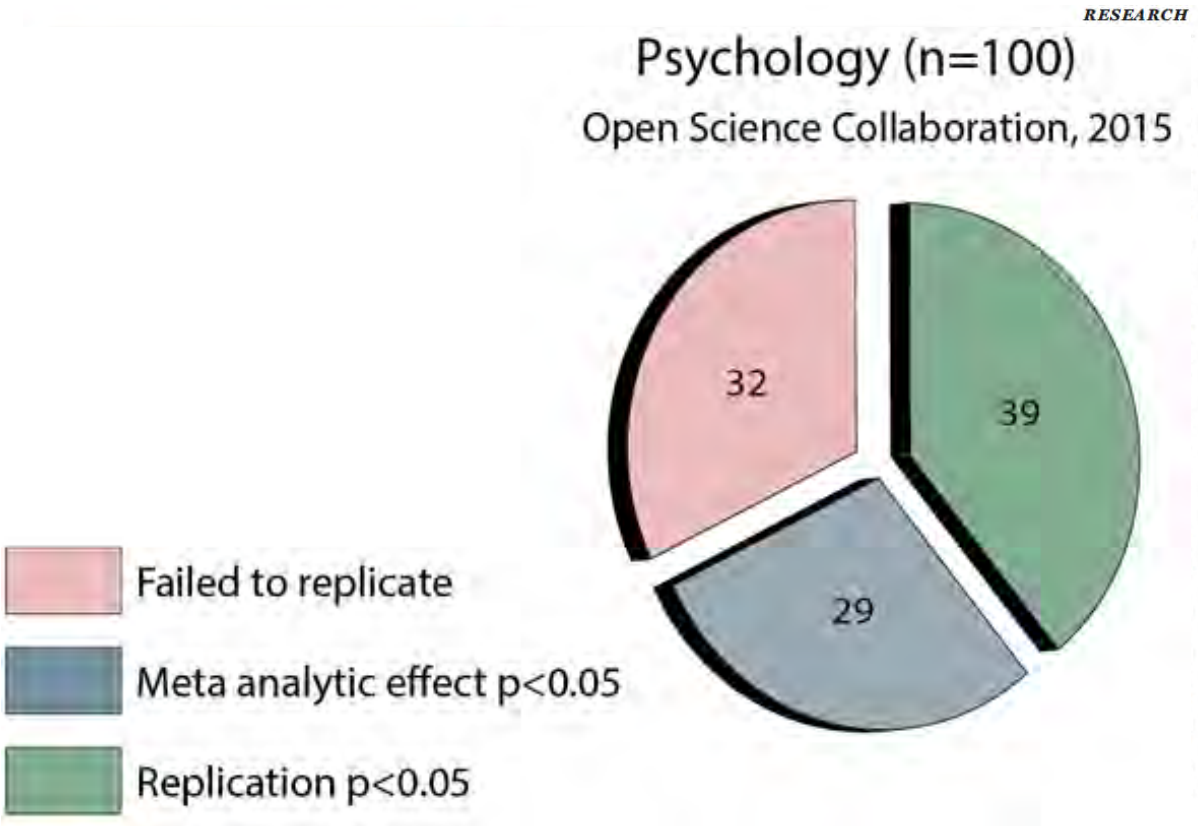
**Conclusions** Contradiction and initially stronger effects are not unusual in highly cited research of clinical interventions and their outcomes. The extent to which high citations may provoke contradictions and vice versa needs more study. Controversies are most common with highly cited nonrandomized studies, but even the most highly cited randomized trials may be challenged and refuted over time, especially small ones.

JAMA. 2005;294:218-228 www.jama.com

CLINICAL RESEARCH ON IMPORTANT questions about the efficacy of medical interventions is sometimes followed by subsequent studies that either reach opposite conclusions or suggest that the original claims were too strong. Such disagreements may upset clinical practice and acquire publicity in both scientific circles and in the lay press. Several empirical investigations have tried to address whether specific types of studies are more likely to be contradicted and to explain observed controversies. For example, evidence exists that small studies may sometimes be refuted by larger ones.[1,2]

Similarly, there is some evidence on disagreements between epidemiological studies and randomized trials.[3-5] Prior investigations have focused on a variety of studies without any particular attention to their relative importance and scientific impact. Yet, most research publications have little impact while a small minority receives most attention and dominates scientific thinking and clinical practice. Im-

# ”Replication crisis”

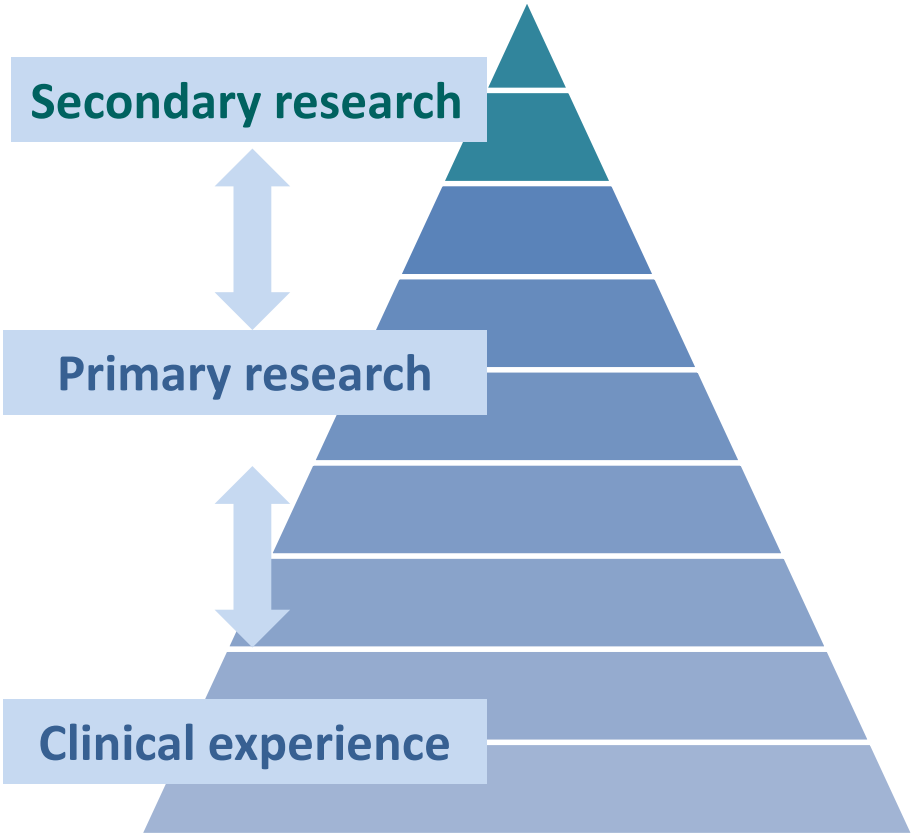



*Collaboration, Open Science (2015-08-28). "Estimating the reproducibility of psychological science". Science.* ***349***

| | Treatment | | | Control | | | |
|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | Sign. |
| Anderson et al. 2000 | 60,3 | 18 | 13 | 68,9 | 16,7 | 12 | ns |
| Burrows et al. 2001 | 57,9 | 11,9 | 20 | 63,7 | 11,3 | 19 | ns |
| Crapper et al. 2002 | 66,5 | 19,5 | 14 | 59,8 | 22,9 | 13 | ns |
| Dim et al. 2003 | 56,9 | 16,7 | 17 | 79,1 | 17,2 | 17 | p < 001 |
| Epstein et al. 2004 | 61,4 | 17,9 | 14 | 65,9 | 16,8 | 14 | ns |
| Fleizig et al. 2005 | 58 | 11,9 | 15 | 61,5 | 11,3 | 15 | ns |

*Is treatment X better than control?*
*(Note: A smaller value is better)*

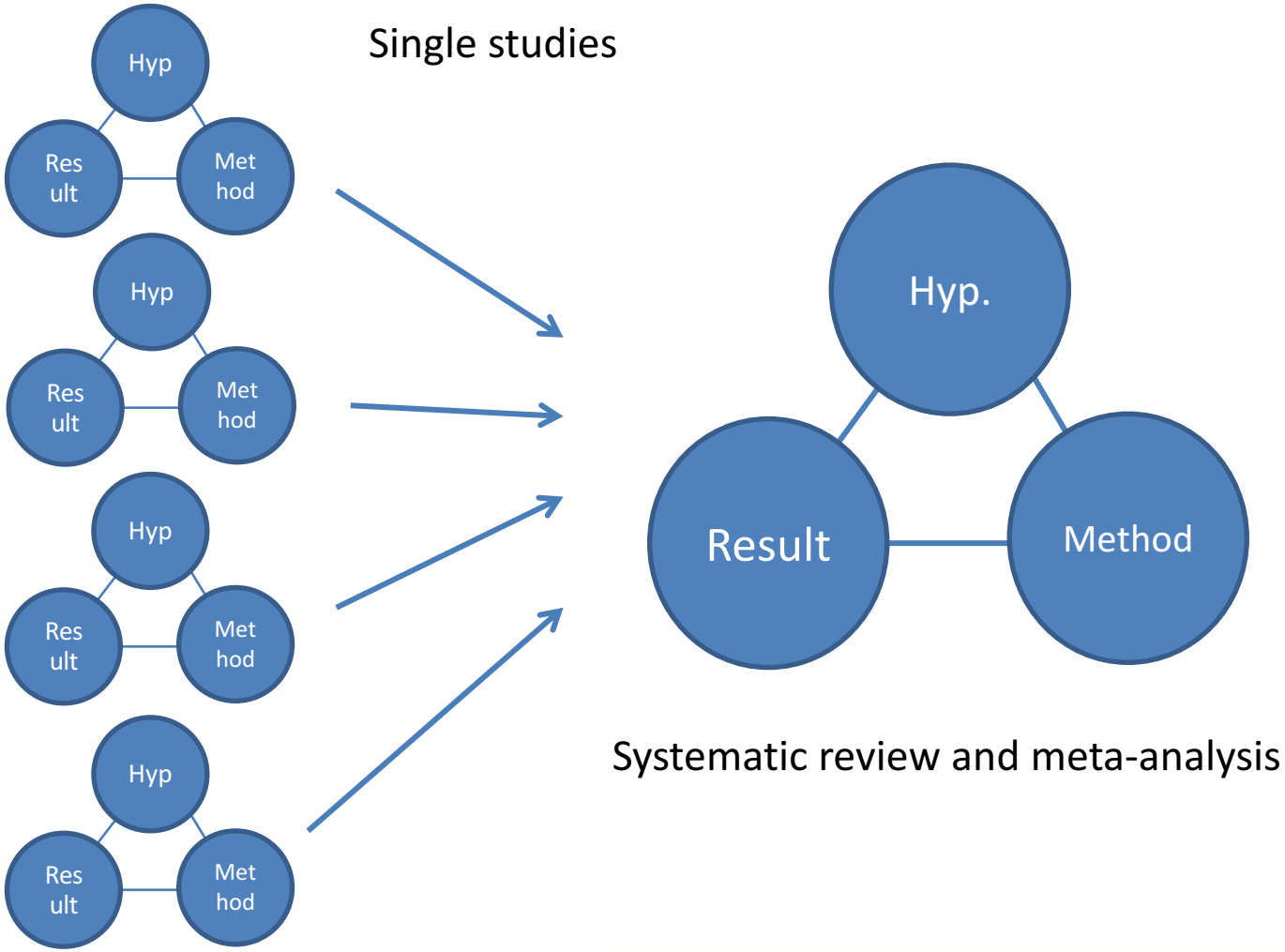
# Evidence hierarchy




- Quantitative systematic review
- Narrative systematic review
- Experimental design (RCT)
- Cohort study
- Case-control study
- Cross sectional study
- Case reports
- Expert opinion
- Anecdote

Challenge: the model favors internal validity

# Systematic reviews and meta-analysis



| | Treatment | | | Control | | | |
|---|---|---|---|---|---|---|---|
| | Mean | SD | N | Mean | SD | N | p |
| Anderson et al. 2000 | 60,3 | 18 | 13 | 68,9 | 16,7 | 12 | 0.22 |
| Burrows et al. 2001 | 57,9 | 11,9 | 20 | 63,7 | 11,3 | 19 | 0.13 |
| Crapper et al. 2002 | 66,5 | 19,5 | 14 | 59,8 | 22,9 | 13 | 0.42 |
| Dim et al. 2003 | 56,9 | 16,7 | 17 | 79,1 | 17,2 | 17 | p < 001 |
| Epstein et al. 2004 | 61,4 | 17,9 | 14 | 65,9 | 16,8 | 14 | 0.50 |
| Fleizig et al. 2005 | 58 | 11,9 | 15 | 61,5 | 11,3 | 15 | 0.41 |

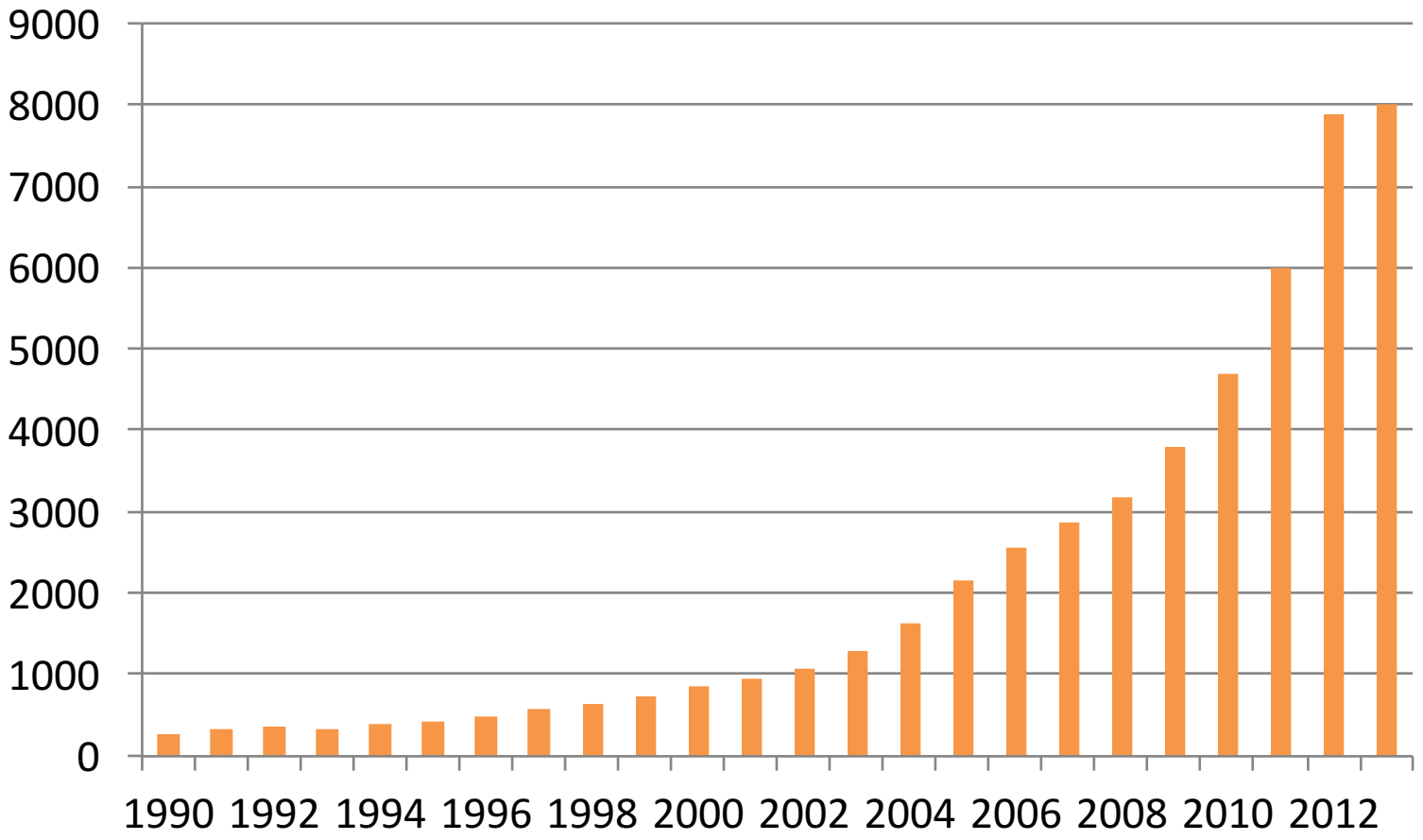
Results of meta-analysis:

| | Sample size | | Heterogeneity [a] | | | | Global effect sizes | | |
|---|---|---|---|---|---|---|---|---|---|
| | K | N | Q | df | *p* | $I^2$ | Hedges *g* | 95% CI | *p* |
| **A. Main effects** | | | | | | | | | |
| Outcome Y (fixed) | 6 | 183 | 9.6 | 5 | 0,09 | 47,7 | 0.420 | 0.13 – 0.71 | **0.004** |

# Systematic reviews and meta-analysis

- Can test (falsify) hypotheses
- Test reproducibility (were initial results random or reliable?)
- Control for random error (variation) between studies
- Test systematic variation between studies
- Generalize results

# Number of published meta-analyses per year - PubMED 1990-2013



# History of meta-analysis

- Pearson (1904) averaged correlations between mortality and inoculation for typhoid fever.
- First medical “meta-analysis” on placebo effects (Beecher,1955)
- Eysenck (1952) argued that psychotherapy was ineffective
- Glass standardized and averaged treatment-control differences from 375 studies, naming it ”meta-analysis” (Smith & Glass, 1977)
- ”An exercise in mega-silliness” (Eysenck, 1978)
- Similar methods developed by Rosenthal and Rubin (1978)
- Today (1977-2014): PubMed: 60460 ”hits” - PsychINFO: 13833 ”hits”
- Cochrane Collaboration (1993): Medicine
- Campbell Collaboration (1999): Social sciences
- Handbook of research synthesis (Cooper & Hedges, 1994)

# Narrative vs Systematic review: A matter of life or death

- From 1972-81, 7 studies investigated the effect of steroid-injektions on premature delivery (associated with increased infant mortality)
- Two studies showed a weak positive effect – the remaining studies were non-significant
- The treatment was abandoned
- A later 1989 meta-analysis of the original data
- revealed a significant positive effect on infant mortality (OR: 0.50)
- The Cochrane Collaboration logo shows data from the 1989 meta-analysis

www.cochrane.org

# Risk of bias


"Excellent health statistics - smokers are less likely to die of age related illness"

# Study quality

- Validity: *"The approximate truth of an inference or claim about a relationship"*
- **Internal validity**
  - Threats: all alternative mechanisms that could explain results, e.g., "placebo", group-differences at baseline, uneven dropout
- **External validity**
  - Are results generalizable to other intended participants and contexts?
- **Construct validity**
  - Do the operational characteristics of intervention and measures adequately represent intended abstract categories?
- **Statistical conclusion validity**
  - The validity of the statistical inferences regarding the strength of the relationship. Threats include insufficient statistical power, regression towards the mean, incorrect assumptions about the underlying variance

Shadish, Cook & Campbell, 2002

# Quality assessment

- All studies should be subjected a pre-defined quality assessment
- Already developed or modified existing checklist
- A newly developed checklist

E.g., Jadad checklist (Jadad, 1996)

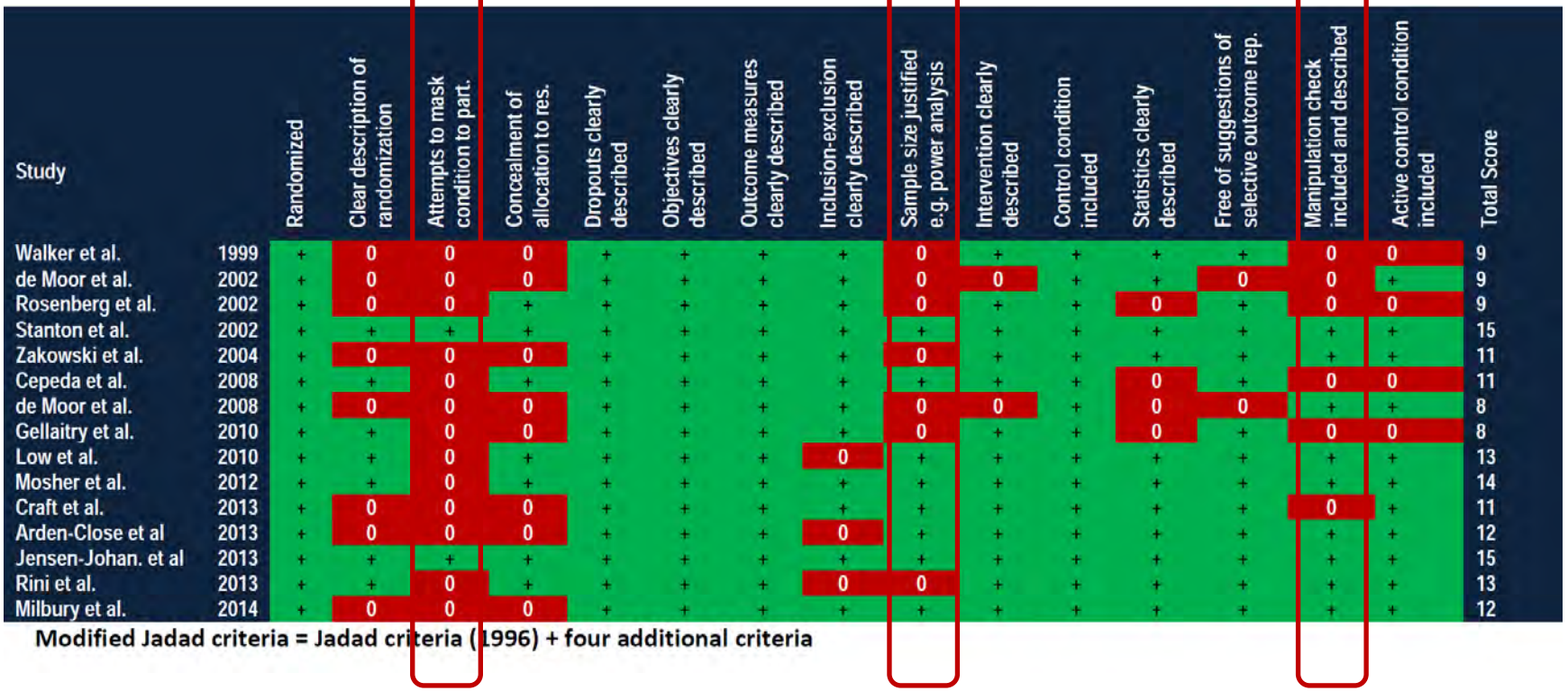
| *Criteria* | score (1, 0, or –1) |
|---|---|
| C1. Did the study include and report both a pre- and post-assessment? *** | |
| C2. Was the study randomized ?* | |
| C3. Was the method of randomization described and appropriate? | |
| C4. Was the randomization described BUT inappropriate? *(deduct 1 pt)* | |
| C5. Was the study described as double blind? | |
| C6. Was any form of blinding of condition (patients) attempted? | |
| C7. Was any form of blinding of assessment/results (researchers) attempted? | |
| C8. Was the blinding described and appropriate (e.g. attention control)? | |
| C9. Was blinding described but inappropriate? *(deduct 1 pt)* | |
| C10. Was there a description of withdrawals and dropouts? | |
| *C11. Were the objectives of the study defined?*** | |
| *C12. Were the outcome measures defined clearly?* | |
| *C13. Was there a clear description of the inclusion and exclusion criteria?* | |
| *C14. Was the sample size justified (e.g. power calculation)?* | |
| *C15. Was there a clear description of the intervention?* | |
| *C16. Was there at least one control (comparison group)?* | |
| *C17. Was the method used to assess adverse effects described?* | |
| *C18. Were the methods of statistical analysis described?* | |
| C19. Were the outcome measure used standardized and reliable? | |
| Jadad 3-item score total (range: 0 – 5) | |
| *Jadad 11-item score total* (range: 0 –13) | |
| Total (modified) quality score (range: 0 – 17) | |

# Quality assessment

Masking conditions

Power analysis

Manipulation check



| Study | | Randomized | Clear description of randomization | Attempts to mask condition to part. | Concealment of allocation to res. | Dropouts clearly described | Objectives clearly described | Outcome measures clearly described | Inclusion-exclusion clearly described | Sample size justified e.g. power analysis | Intervention clearly described | Control condition included | Statistics clearly described | Free of suggestions of selective outcome rep. | Manipulation check included and described | Active control condition included | Total Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Walker et al. | 1999 | + | 0 | 0 | 0 | + | + | + | + | 0 | + | + | + | + | 0 | 0 | 9 |
| de Moor et al. | 2002 | + | 0 | 0 | 0 | + | + | + | + | 0 | 0 | + | + | 0 | 0 | + | 9 |
| Rosenberg et al. | 2002 | + | 0 | 0 | + | + | + | + | + | 0 | + | + | 0 | + | 0 | 0 | 9 |
| Stanton et al. | 2002 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | 15 |
| Zakowski et al. | 2004 | + | 0 | 0 | 0 | + | + | + | + | 0 | + | + | + | + | + | + | 11 |
| Cepeda et al. | 2008 | + | + | 0 | + | + | + | + | + | + | + | + | 0 | + | 0 | 0 | 11 |
| de Moor et al. | 2008 | + | 0 | 0 | 0 | + | + | + | + | 0 | 0 | + | 0 | 0 | + | + | 8 |
| Gellaitry et al. | 2010 | + | + | 0 | 0 | + | + | + | + | 0 | + | + | 0 | + | 0 | 0 | 8 |
| Low et al. | 2010 | + | + | 0 | + | + | + | + | 0 | + | + | + | + | + | + | + | 13 |
| Mosher et al. | 2012 | + | + | 0 | + | + | + | + | + | + | + | + | + | + | + | + | 14 |
| Craft et al. | 2013 | + | 0 | 0 | 0 | + | + | + | + | + | + | + | + | + | 0 | + | 11 |
| Arden-Close et al | 2013 | + | 0 | 0 | 0 | + | + | + | 0 | + | + | + | + | + | + | + | 12 |
| Jensen-Johan. et al | 2013 | + | + | + | + | + | + | + | + | + | + | + | + | + | + | + | 15 |
| Rini et al. | 2013 | + | + | 0 | + | + | + | + | 0 | 0 | + | + | + | + | + | + | 13 |
| Milbury et al. | 2014 | + | 0 | 0 | 0 | + | + | + | + | + | + | + | + | + | + | + | 12 |

Modified Jadad criteria = Jadad criteria (1996) + four additional criteria

Possible score: 0-15

Mean score = 11.3 (SD = 2.4; range: 8-15)

Inter-rater agreement: 89.9% of 225 individual quality ratings

Zachariae & O'Toole, 2015

# ES of published vs unpublished

**Table 3**
*Comparison of Effect Sizes Reported in Published Versus Unpublished Studies*

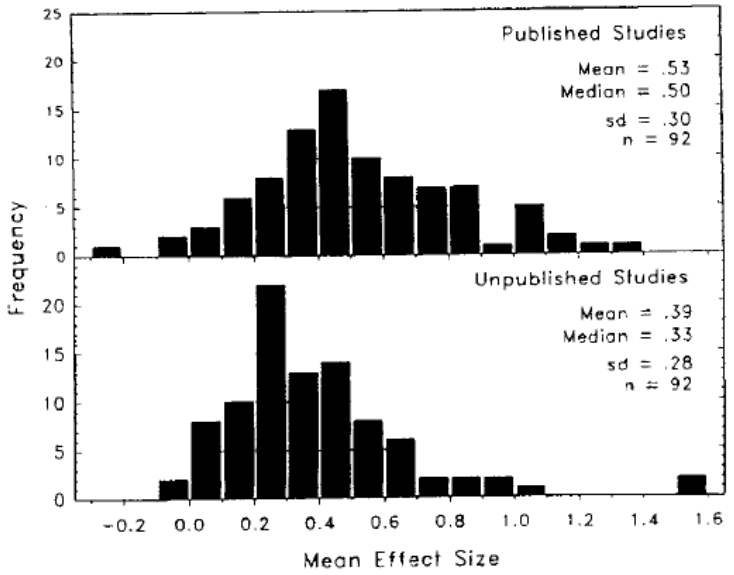
| Document source | Effect size M | Effect size SD | N |
|---|---|---|---|
| Published studies | 0.53 | 0.30 | 92 |
| Unpublished studies | 0.39 | 0.28 | 92 |

*Note.* Only those meta-analyses that provided a breakout for this construct were included.

timate derived from published studies and that derived from unpublished studies within the same set of meta-analyses. Published studies yielded mean effect sizes that averaged 0.14 *SD*s larger than unpublished studies. It is evident, therefore, that the treatment effects reported in published studies are indeed generally biased upward, relative to those in unpublished studies.

It is noteworthy, however, that the mean effect size estimates for both published and unpublished studies fall in the positive range; published studies are just more pos-

**Figure 5**
*Distributions of Mean Effect Sizes From Published and Unpublished Studies for Meta-Analyses Reporting Both Breakouts*

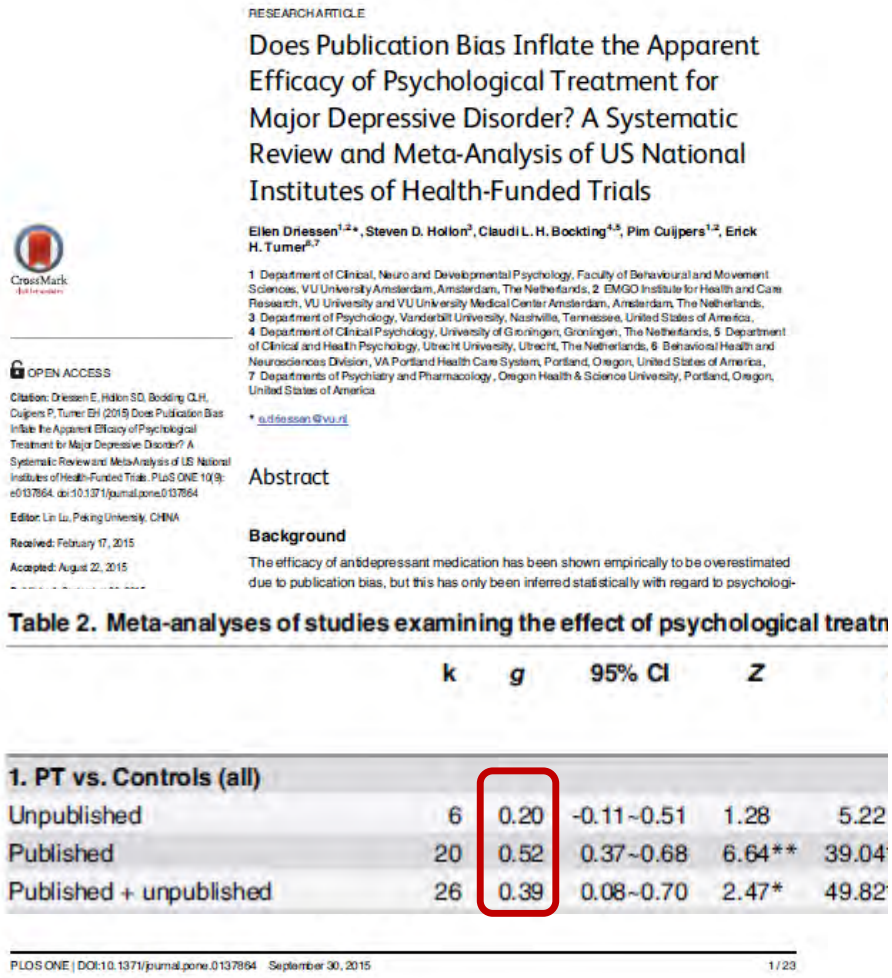



Lipsey & Wilson, 1993

# Examples of publication bias

- Medical journals from China almost never publish negative results (e.g. Pan et al. 2005)
- Only 5% of articles in journals focusing on Alternative and Complementary Medicine present negative results (Schmidt et al. 2001)
- Studies originating from Europe have more positive results than studies from the US (Sood et al. 2007)

# Publication bias



RESEARCH ARTICLE

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

Ellen Driessen[1,2]*, Steven D. Hollon[3], Claudi L. H. Bockting[4,5], Pim Cuijpers[1,2], Erick H. Turner[6,7]

1 Department of Clinical, Neuro and Developmental Psychology, Faculty of Behavioural and Movement Sciences, VU University Amsterdam, Amsterdam, The Netherlands, 2 EMGO Institute for Health and Care Research, VU University and VU University Medical Center Amsterdam, Amsterdam, The Netherlands, 3 Department of Psychology, Vanderbilt University, Nashville, Tennessee, United States of America, 4 Department of Clinical Psychology, University of Groningen, Groningen, The Netherlands, 5 Department of Clinical and Health Psychology, Utrecht University, Utrecht, The Netherlands, 6 Behavioral Health and Neurosciences Division, VA Portland Health Care System, Portland, Oregon, United States of America, 7 Departments of Psychiatry and Pharmacology, Oregon Health & Science University, Portland, Oregon, United States of America

* e.driessen@vu.nl

OPEN ACCESS

Citation: Driessen E, Hollon SD, Bockting CLH, Cuijpers P, Turner EH (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(9): e0137864. doi:10.1371/journal.pone.0137864

Editor: Lin Lu, Peking University, CHINA

Received: February 17, 2015

Accepted: August 22, 2015

Abstract

Background

The efficacy of antidepressant medication has been shown empirically to be overestimated due to publication bias, but this has only been inferred statistically with regard to psychologi-

**Table 2. Meta-analyses of studies examining the effect of psychological treatn**

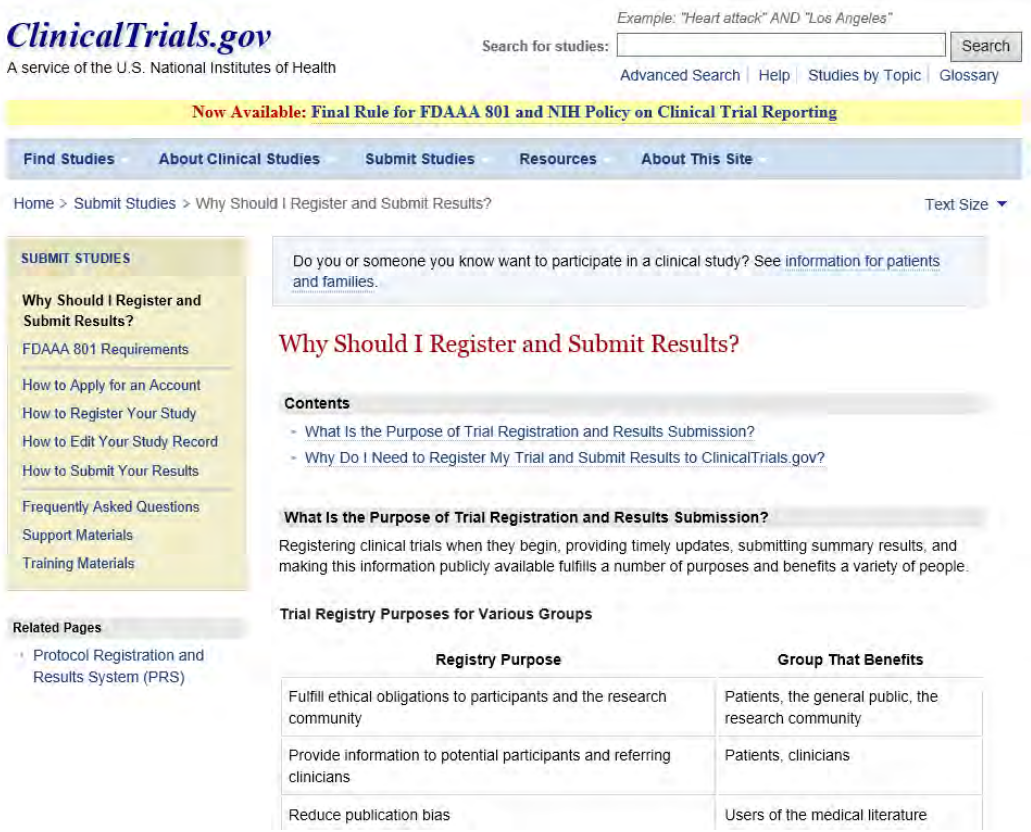
| | k | g | 95% CI | z | |
|---|---|---|---|---|---|
| **1. PT vs. Controls (all)** | | | | | |
| Unpublished | 6 | 0.20 | -0.11~0.51 | 1.28 | 5.22 |
| Published | 20 | 0.52 | 0.37~0.68 | 6.64** | 39.04 |
| Published + unpublished | 26 | 0.39 | 0.08~0.70 | 2.47* | 49.82 |



- CONCLUSION:
- The efficacy of psychological interventions for depression has been overestimated in the published literature
- Just as it has been for pharmacotherapy.
- Both are efficacious but not to the extent that the published literature would suggest.

# Preregistering

- ClinicalTrials.gov
- Aims:
  - Increase transparency
  - Reduce fishing expedition bias
  - Presentation of *post-hoc* hypotheses as *a priory*
  - Enable assessment of publication bias



ClinicalTrials.gov
A service of the U.S. National Institutes of Health

Example: "Heart attack" AND "Los Angeles"
Search for studies: Search
Advanced Search | Help | Studies by Topic | Glossary

Now Available: Final Rule for FDAAA 801 and NIH Policy on Clinical Trial Reporting

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

Home > Submit Studies > Why Should I Register and Submit Results?

Text Size

SUBMIT STUDIES

Why Should I Register and Submit Results?
FDAAA 801 Requirements
How to Apply for an Account
How to Register Your Study
How to Edit Your Study Record
How to Submit Your Results
Frequently Asked Questions
Support Materials
Training Materials

Related Pages
Protocol Registration and Results System (PRS)

Do you or someone you know want to participate in a clinical study? See information for patients and families.

## Why Should I Register and Submit Results?

Contents
- What Is the Purpose of Trial Registration and Results Submission?
- Why Do I Need to Register My Trial and Submit Results to ClinicalTrials.gov?

### What Is the Purpose of Trial Registration and Results Submission?

Registering clinical trials when they begin, providing timely updates, submitting summary results, and making this information publicly available fulfills a number of purposes and benefits a variety of people.

Trial Registry Purposes for Various Groups

| Registry Purpose | Group That Benefits |
|---|---|
| Fulfill ethical obligations to participants and the research community | Patients, the general public, the research community |
| Provide information to potential participants and referring clinicians | Patients, clinicians |
| Reduce publication bias | Users of the medical literature |

# Publication bias assessment

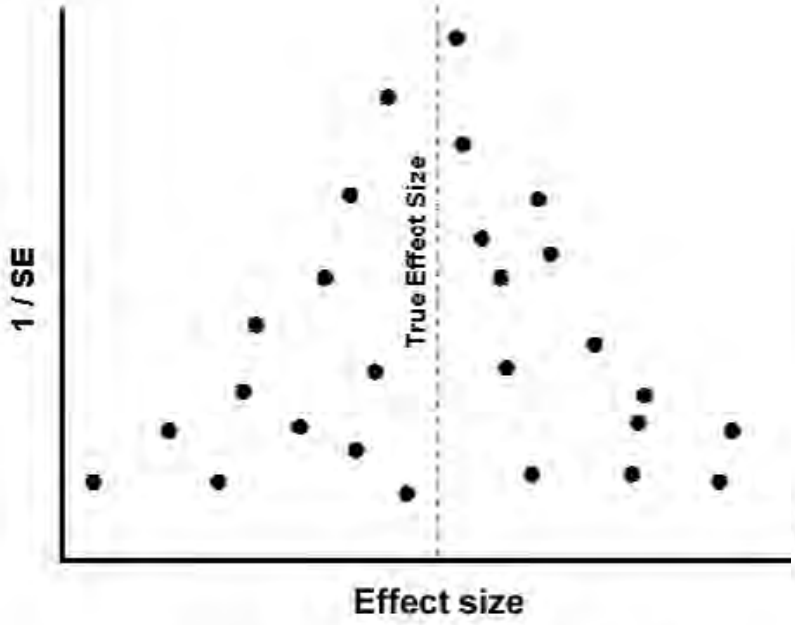



Figure 1.
A fictitious funnel plot with no publication bias

Researcher bias against submitting negative results
Publisher bias against publishing negative results

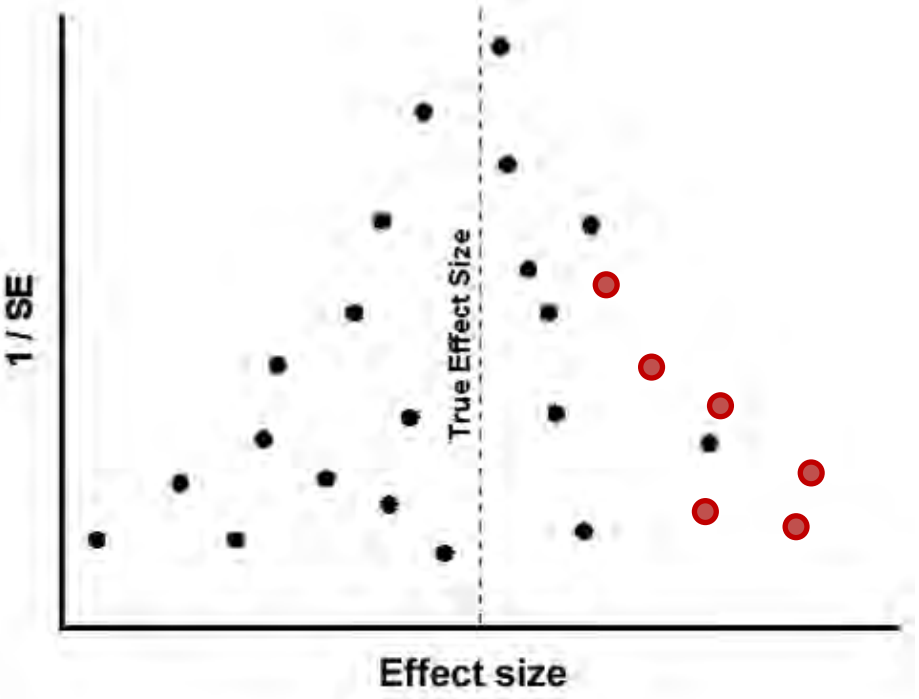


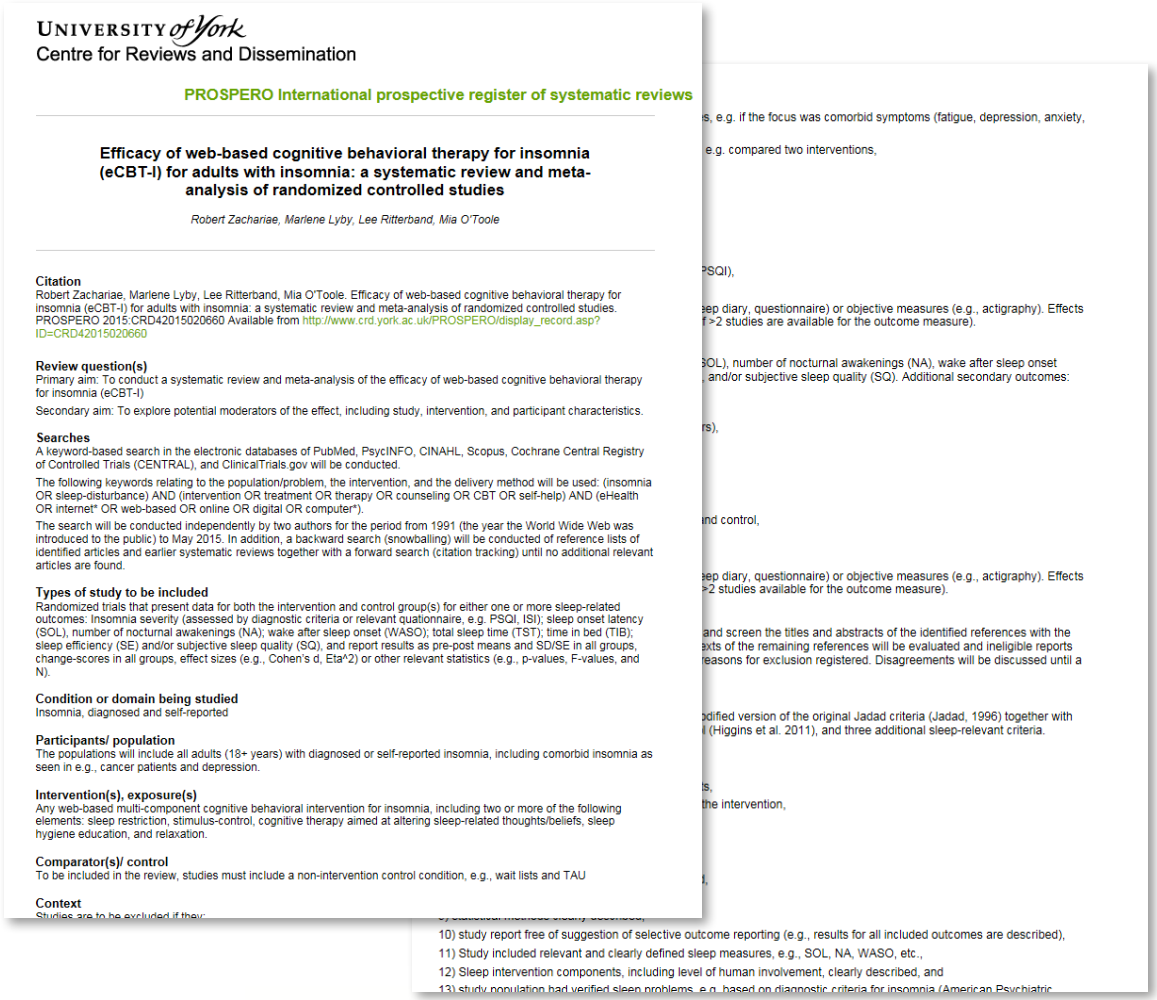

Figure 2
The same plot showing publication bias.

**Cooper, H., DeNeve, K. & Charlton, K. (1997).**

**Begg, C.B. & Berlin, J.A. (1988).**

# Preregistering

- PROSPERO
  - Centre for Reviews and Dissemination



UNIVERSITY of York
Centre for Reviews and Dissemination

PROSPERO International prospective register of systematic reviews

**Efficacy of web-based cognitive behavioral therapy for insomnia (eCBT-I) for adults with insomnia: a systematic review and meta-analysis of randomized controlled studies**

*Robert Zachariae, Marlene Lyby, Lee Ritterband, Mia O'Toole*

**Citation**
Robert Zachariae, Marlene Lyby, Lee Ritterband, Mia O'Toole. Efficacy of web-based cognitive behavioral therapy for insomnia (eCBT-I) for adults with insomnia: a systematic review and meta-analysis of randomized controlled studies. PROSPERO 2015:CRD42015020660 Available from http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42015020660

**Review question(s)**
Primary aim: To conduct a systematic review and meta-analysis of the efficacy of web-based cognitive behavioral therapy for insomnia (eCBT-I)

Secondary aim: To explore potential moderators of the effect, including study, intervention, and participant characteristics.

**Searches**
A keyword-based search in the electronic databases of PubMed, PsycINFO, CINAHL, Scopus, Cochrane Central Registry of Controlled Trials (CENTRAL), and ClinicalTrials.gov will be conducted.

The following keywords relating to the population/problem, the intervention, and the delivery method will be used: (insomnia OR sleep-disturbance) AND (intervention OR treatment OR therapy OR counseling OR CBT OR self-help) AND (eHealth OR internet* OR web-based OR online OR digital OR computer*).

The search will be conducted independently by two authors for the period from 1991 (the year the World Wide Web was introduced to the public) to May 2015. In addition, a backward search (snowballing) will be conducted of reference lists of identified articles and earlier systematic reviews together with a forward search (citation tracking) until no additional relevant articles are found.

**Types of study to be included**
Randomized trials that present data for both the intervention and control group(s) for either one or more sleep-related outcomes: Insomnia severity (assessed by diagnostic criteria or relevant quationnaire, e.g. PSQI, ISI); sleep onset latency (SOL), number of nocturnal awakenings (NA); wake after sleep onset (WASO); total sleep time (TST); time in bed (TIB); sleep efficiency (SE) and/or subjective sleep quality (SQ), and report results as pre-post means and SD/SE in all groups, change-scores in all groups, effect sizes (e.g., Cohen's d, Eta^2) or other relevant statistics (e.g., p-values, F-values, and N).

**Condition or domain being studied**
Insomnia, diagnosed and self-reported

**Participants/ population**
The populations will include all adults (18+ years) with diagnosed or self-reported insomnia, including comorbid insomnia as seen in e.g., cancer patients and depression.

**Intervention(s), exposure(s)**
Any web-based multi-component cognitive behavioral intervention for insomnia, including two or more of the following elements: sleep restriction, stimulus-control, cognitive therapy aimed at altering sleep-related thoughts/beliefs, sleep hygiene education, and relaxation.

**Comparator(s)/ control**
To be included in the review, studies must include a non-intervention control condition, e.g., wait lists and TAU

**Context**
Studies are to be excluded if they:

10) study report free of suggestion of selective outcome reporting (e.g., results for all included outcomes are described),
11) Study included relevant and clearly defined sleep measures, e.g., SOL, NA, WASO, etc.,
12) Sleep intervention components, including level of human involvement, clearly described, and

# Examples

**The Concept of a Systematic Review**


Unit for Psychooncology and Health Psychology
AARHUS UNIVERSITY
Aarhus University Hospital
Robert Zachariae
Professor, MSc. DMSci
bzach@aarhus.rm.dk

# Internet-delivered CBT for insomnia

- Annual prevalence of insomnia: 10-20%
- 6% with a chronic trajectory
- Pharmacological treatment is non-curative and long-term use is associated with dependence, tolerance, side-effects, and increased mortality
- Cognitive Behavioral Therapy for Insomnia (CBT-I) is recommended as first choice – based on evidence from systematic reviews and meta-analyses
- Limited availability of CBT-I (trained therapists, geography, financial reasons)
- One possibility is Internet-delivered CBT-I (eCBT-I)
- *Is eCBT-I effective and are effects comparable to face-to-face delivered CBT-I?*

Vidensråd for forebyggelse (2015) Søvn og sundhed; American Academy of Sleep Medicine

# Internet-delivered CBT-I

Contents lists available at ScienceDirect

Sleep Medicine Reviews

journal homepage: www.elsevier.com/locate/smrv

CLINICAL REVIEW

## Efficacy of internet-delivered cognitive-behavioral therapy for insomnia – A systematic review and meta-analysis of randomized controlled trials

Robert Zachariae [a, *], Marlene S. Lyby [a], Lee M. Ritterband [b], Mia S. O'Toole [a]

[a] Unit for Psychooncology and Health Psychology, Department of Oncology, Aarhus University Hospital, and Department of Psychology and Behavioural Science, Aarhus University, Aarhus, Denmark
[b] Behavioral Health and Technology Lab, Dept. of Psychiatry and Neurobehavioral Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA

ARTICLE INFO

*Article history:*
Received 28 July 2015
Received in revised form
4 October 2015
Accepted 14 October 2015
Available online 24 October 2015

*Keywords:*
Insomnia
internet-delivered
Cognitive-behavioral therapy
eHealth

SUMMARY

Cognitive-behavioral therapy for insomnia (CBT-I) has been shown efficacious, but the challenge remains to make it available and accessible in order to meet population needs. Delivering CBT-I over the internet (eCBT-I) may be one method to overcome this challenge. The objective of this meta-analysis was to evaluate the efficacy of eCBT-I and the moderating influence of various study characteristics. Two researchers independently searched key electronic databases (1991 to June 2015), selected eligible publications, extracted data, and evaluated methodological quality. Eleven randomized controlled trials examining a total of 1460 participants were included. Results showed that eCBT-I improved insomnia severity, sleep efficiency, subjective sleep quality, wake after sleep onset, sleep onset latency, total sleep time, and number of nocturnal awakenings at post-treatment, with effect sizes (Hedges's *g*) ranging from 0.21 to 1.09. The effects were comparable to those found for face-to-face CBT-I, and were generally maintained at 4–48 wk follow-up. Moderator analyses showed that longer treatment duration and higher degree of personal clinical support were associated with larger effect sizes, and that larger study dropout in the intervention group was associated with smaller effect sizes. In conclusion, internet-delivered CBT-I appears efficacious and can be considered a viable option in the treatment of insomnia.


### Introduction

Insomnia is a common clinical condition with an annual prevalence of 10–20% [1] and approximately half (6%) with a chronic trajectory, i.e., persisting for more than 3 mo [2]. Chronic insomnia has been associated with a number of negative health outcomes, including obesity and metabolic dysregulation [3,4], hypertension and increased risk of myocardial infarction [5,6], increased susceptibility to infections [7], and depression [8]. While pharmacotherapy remains the most commonly used treatment option, hypnotics such as benzodiazepine receptor agonists are associated with side-effects, dependence, and tolerance over time. They are usually not curative, leading to long-term treatment over many years despite lack of safety and efficacy data beyond 1–2 y [9,10]. In contrast, cognitive-behavioral therapy for insomnia (CBT-I) has, in several meta-analyses, been found efficacious in improving sleep outcomes [11–13] with acute effects comparable or superior to those found for pharmacotherapy [14], and these effects have been maintained for up to 3 y [10]. While CBT-I has been shown efficacious and desired by many patients preferring non-pharmacological approaches [15,16], the challenge remains to make it available and accessible to meet population needs [17] due to the limited availability of trained therapists and the relatively high costs of CBT-I delivered face-to-face [18,19].

One method to overcome these challenges may be to provide CBT-I over the internet [20]. The first randomized controlled trial (RCT) evaluating internet-delivered CBT-I, published in 2004, compared a 5-wk internet-delivered sleep management program to a waiting list control in a sample with clinical insomnia [21]. The program, which was derived from existing CBT-I manuals [22,23],

* Corresponding author. Unit for Psychooncology and Health Psychology, Dept. of Oncology, Aarhus University Hospital and Dept. of Psychology and Behavioural Science, Aarhus University, Bartholins Allé 9, Bld. 1340, 8000 Aarhus, Denmark. Tel.: +45 87162674.
*E-mail address:* bzach@aarhus.rm.dk (R. Zachariae).

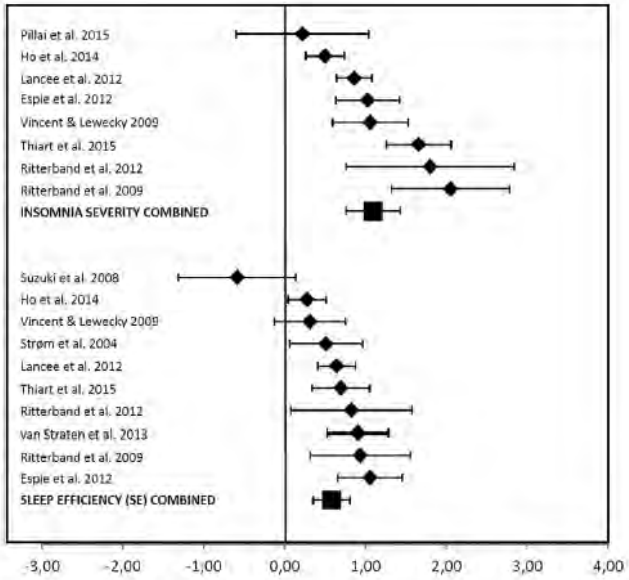



Fig. 2. Forest plot of post-treatment effect sizes for insomnia severity and sleep efficiency (SE).

Statistically significant effects found for primary outcomes:

Insomnia severity:

Hedges's g = 1.09, $p < 0.001$

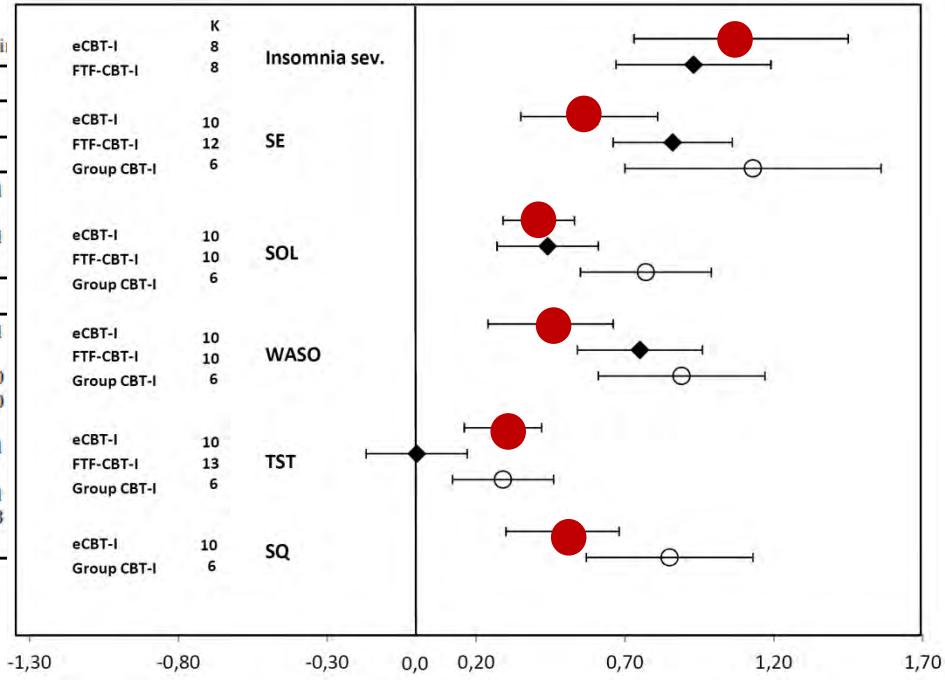
Sleep efficiency:

Hedges's g = 0.58, $p < 0.001$

# Internet-delivered CBT-I

**Table 2**
Pooled post-treatment effects of internet-delivered cognitive behavioral therapy for i

| Outcome | Sample size | | Heterogeneity[b] | | |
|---|---|---|---|---|---|
| | K | N[a] | Q | df | *p* |
| **Primary outcomes** | | | | | |
| Insomnia severity[f] | 8 | 1071 | 40.7 | 7 | <0.001 |
| *Adjusted for publication bias*[g] | *(9)* | | | | |
| Sleep efficiency (SE) | 10 | 1220 | 28.5 | 9 | 0.001 |
| *Adjusted for publication bias* | *(11)* | | | | |
| **Secondary outcomes** | | | | | |
| Sleep onset latency (SOL) | 10 | 1114 | 8.6 | 9 | 0.471 |
| *Adjusted for publication bias* | *(14)* | | | | |
| Wake after sleep onset (WASO) | 7 | 944 | 11.6 | 6 | 0.070 |
| No. of nocturnal awakenings (NA) | 6 | 640 | 5.1 | 5 | 0.400 |
| *Adjusted for publication bias* | *(9)* | | | | |
| Total sleep time (TST) | 10 | 1114 | 9.5 | 9 | 0.391 |
| *Adjusted for publication bias* | *(12)* | | | | |
| Time in bed (TIB) | 4 | 464 | 1.5 | 3 | 0.681 |
| Subjective sleep quality (SQ) | 8 | 801 | 10.7 | 7 | 0.153 |
| *Adjusted for publication bias* | *(10)* | | | | |




Comparing with face-to-face delivery

No statistically significant differences between internet-delivered and face-to-face-delivered CBT-I. Need for *non-inferiority trials* directly comparing eCBT-I and FtF

# Psychological intervention for distress in informal cancer caregivers

- Informal cancer caregivers (ICCs) report increased levels of psychological and physical morbidity and higher mortality
- Psychological interventions such as Cognitive Behavioral Therapies (CBTs) have been shown efficacious for distress (anxiety and depression)
- Aim: *to evaluate the available evidence for the efficacy of CBTs for distress and physical symptoms among ICCs*

# Psychological intervention for distress in informal cancer caregivers

**Psycho-Oncology**
*Psycho-Oncology* (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI**: 10.1002/pon.4144

**Review**

## Cognitive behavioral therapies for informal caregivers of patients with cancer and cancer survivors: a systematic review and meta-analysis

Mia S. O'Toole[1]*, Robert Zachariae[2], Megan E. Renna[3], Douglas S. Mennin[4] and Allison Applebaum[5]
[1]Unit for Psychooncology and Health Psychology, Institute of Psychology and Behavioral Science, School of Business and Social Sciences, Aarhus University, Aarhus, Denmark
[2]Unit for Psychooncology and Health Psychology, Aarhus University and Aarhus University Hospital, Aarhus, Denmark
[3]Department of Psychology, CUNY Hunter College and the Graduate Center, New York, USA
[4]Department of Psychology, CUNY Hunter College and the Graduate Center, New York, USA
[5]Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA

*Correspondence to: Unit for Psychooncology and Health Psychology, Institute of Psychology and Behavioral Science, School of Business and Social Sciences, Bartholins Allé 9, 8000 Aarhus C, Denmark. E-mail: mia@psy.au.dk

Received: 2 December 2015
Revised: 11 March 2016
Accepted: 25 March 2016

**Abstract**

***Objective*: Informal caregivers (ICs) of patients with cancer and cancer survivors report a number of psychological and physical complaints because of the burden associated with providing care. Given the documented effect of Cognitive Behavioral Therapy (CBT) on ICs' common psychological complaints, such as anxiety and depression, the objective was to conduct a meta-analysis on the effect of CBTs for adult ICs.**

***Methods*: A literature search was conducted in order to identify all intervention studies on adult ICs that employed at least one therapeutic component defined as a CBT component.**

***Results*: Literature searches revealed 36 unique records with sufficient data. These studies were subjected to meta-analyses using random effects models. A small, statistically significant effect of CBTs (Hedge's $g = 0.08$, $p = 0.014$) was revealed, which disappeared when randomized controlled trials were evaluated alone ($g = 0.04$, $p = 0.200$). A number of variables were explored as moderators. Only the percentage of female participants was positively associated with the effect size.**

***Conclusions*: Based on the negligible effect of CBTs across outcomes, future studies should consider moving beyond traditional CBT methods as these do not appear efficacious. It is suggested that future interventions orient towards advances in the basic affective sciences and derived therapies in order to better understand and treat the emotional struggles experienced by ICs.**



### Background

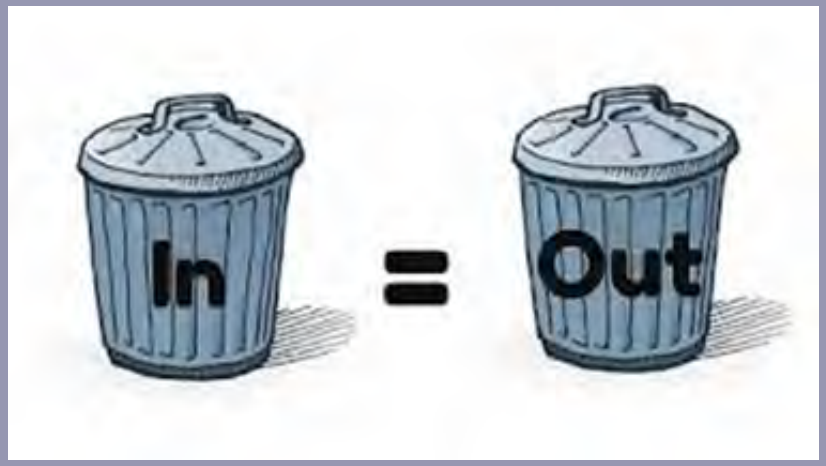
There is growing recognition that informal caregivers (ICs) of chronically ill patients are themselves in need of care. Historically, research on caregiver burden has focused on ICs of patients with a variety of dementias, such as Alzheimer's and Parkinson's disease. More recently, the burden experienced by ICs of patients with cancer is receiving increased attention, which may in part be because of the rising incidence of cancer globally [1]. Such psychological and physical complaints [3,4]. ICs have been found to have high levels of psychological distress, and longitudinal studies have shown that caregiver burden is significantly associated with anxiety and depression over time [e.g. 5,6]. Examples of specific complaints by ICs include feeling overwhelmed by taking on the patient's responsibilities, fear of losing the patient, and uncertainty about the future [7]. Such complaints are likely to persist into survivorship, as 30–40% of caregivers continue to experience clinical

**Table 2.** Results from meta-regression-based moderation analyses

| Variable | Unadjusted model[a] | | | Adjusted model[b] | | | |
|---|---|---|---|---|---|---|---|
| | *B* | SE | p | *B* | SE | p | *K* |
| *Study quality characteristics*[c] | | | | | | | |
| Design (RCT vs. OT) | −0.17 | 0.08 | **0.028** | | | | |
| Control type (Active vs. non-active) | −0.15 | 0.08 | 0.059 | | | | |
| JADAD | −0.00 | 0.02 | 0.874 | | | | |
| *Caregiver characteristics* | | | | | | | 29 |
| Age | −0.01 | 0.01 | 0.468 | −0.02 | 0.01 | **0.046** | |
| % women | <0.01 | <0.01 | **0.002** | 0.01 | <0.01 | **0.001** | |
| Interaction[d] | | | | −0.00 | <0.01 | 0.731 | |
| *Intervention characteristics* | | | | | | | 24 |
| # sessions | −0.01 | 0.01 | 0.484 | 0.01 | 0.05 | 0.809 | |
| Treatment duration | −0.01 | <0.01 | 0.154 | −0.02 | 0.02 | 0.310 | |
| # components | 0.02 | 0.02 | 0.454 | −0.06 | 0.08 | 0.451 | |
| CBT (CBT vs. other) | 0.01 | 0.08 | 0.855 | 0.11 | 0.18 | 0.541 | |
| Recipient (IC vs. group/dyad) | 0.05 | 0.09 | 0.541 | 0.06 | 0.12 | 0.578 | |
| Modality (face-to-face vs. web/phone) | 0.04 | 0.08 | 0.639 | −0.03 | 0.17 | 0.869 | |
| Format (individual vs. group) | −0.09 | 0.09 | 0.313 | −0.01 | 0.20 | 0.953 | |
| *Patient characteristics* | | | | | | | 13 |
| Stage | | | | | | | |
| Mixed (vs. early) | 0.01 | 0.12 | 0.939 | −0.14 | 0.21 | 0.503 | |
| Late (vs. early) | −0.05 | 0.12 | 0.648 | −0.23 | 0.20 | 0.252 | |
| Survivor (vs. early) | 0.21 | 0.18 | 0.247 | −0.25 | 0.33 | 0.450 | |
| Time since diagnosis | 0.02 | 0.04 | 0.610 | 0.04 | 0.04 | 0.431 | |

O'Toole, Zachariae, Penna, Mennin, Applebaum, 2016

# Criticisms of meta-analysis



# Criticisms of meta-analysis

*"Exercise in mega-silliness"* (H. Eysenck, 1978) *"Statistical alchemy"* (Feinstein, 1995)

- Mixes apples and oranges

- Garbage in – garbage out


- File-drawer problem

- Reductionism (one number)

***d* = 2,7 (95%CI: 1,7 – 3,7)**

# Systematic reviews

| | Transparency | Risk of bias | Effect estimation |
|---|---|---|---|
| Non-systematic review | **Low** | **High** | **None** |
| Narrative systematic review | **Medium-high** | **Medium** | **Qualitative**<br>**"Vote counting"** |
| Quantitative systematic review (meta-analysis) | **High** | **Low** | **Magnitude**<br>**Direction**<br>**Precision**<br>**Sub-group comparisons** |

# Lessons learned

*”One must seek the truth where it is, not where one would like it to be”*

*Abbé de Faria, 1746-1819*

# Lessons learned

- When evaluating evidence
  - One study is not enough (replicability)
  - Avoid cherry picking – focus on the combined evidence of all available evidence
  - Less emphasis on p-values of individual studies – more emphasis on magnitude (effect size)
  - More emphasis on practical significance (e.g., MCID)
    - A highly statistically significant effect could be of a irrelevant magnitude
    - A non-statistically significant effect could potentially be clinically relevant

# Lessons learned

- Consider
  - Statistical power and risk of Type-2 error
  - The precision of the effect (the confidence interval)
  - Homogeneity of the existing evidence
  - Study quality – potential bias and threats to validity
  - Publication bias – the tendency to underreport null-findings
  - Cost-effectiveness – relative to treatment as usual
- Establishing evidence
  - Is a complex cumulative process

# Future tasks

- The future will bring public and policy-based demands for evaluation of clinical efficacy and cost-effectiveness
- Psychologists are advised to:
  - Work to establish a proactive, evidence-based professional culture
  - Not to take the effectiveness of psychological approaches as self-evident but to focus on the best available evidence
  - Accept when psychological approaches are not effective, be transparent about it, and work to improve the situation
  - Promote research-based practice and practice-relevant research, establish collaboration between researchers and clinicians, and conduct research-based evaluation in collaborative networks

Patient Expectancy and Post-chemotherapy Nausea:

Contents lists available at ScienceDirect

Influence of Psychological Stress on Upper Respiratory Infection—A Meta-Ar

ANETTE

ELSEVIER

CLINICAL REVIEW

Efficacy of inter
insomnia – A s
controlled trials

Robert Zachariae [a,

[a] Unit for Psychooncology and He
Science, Aarhus University, Aarhu
[b] Behavioral Health and Technolo
USA

ARTICLE INFO

*Article history:*
Received 28 July 2015
Received in revised form
4 October 2015
Accepted 14 October 2015
Available online 24 October 201

*Keywords:*
Insomnia
internet-delivered
Cognitive-behavioral therapy
eHealth

**Psycho-Oncology**
*Psycho-Oncology* (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com). **DOI**: 10.1002/pon.4144

**Review**

# Cognitive behavioral therapies for informal caregivers of patients with cancer and cancer survivors: a systematic review and meta-analysis

Mia S. O'Toole[1]*, Robert Zachariae[2], Megan E. Renna[3], Douglas S. Mennin[4] and Allison Applebaum[5]
[1]*Unit for Psychooncology and Health Psychology, Institute of Psychology and Behavioral Science, School of Business and Social Sciences, Aarhus University, Aarhus, Denmark*
[2]*Unit for Psychooncology and Health Psychology, Aarhus University and Aarhus University Hospital, Aarhus, Denmark*
[3]*Department of Psychology, CUNY Hunter College and the Graduate Center, New York, USA*
[4]*Department of Psychology, CUNY Hunter College and the Graduate Center, New York, USA*
[5]*Department of Psychiatry and Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, USA*

*Correspondence to:
*Unit for Psychooncology and Health Psychology, Institute of Psychology and Behavioral Science, School of Business and Social Sciences, Bartholins Allé 9, 8000 Aarhus C, Denmark.
*E-mail: mia@psy.au.dk*

## Abstract

***Objective***: Informal caregivers (ICs) of patients with cancer and cancer survivors report a number of psychological and physical complaints because of the burden associated with providing care. Given the documented effect of Cognitive Behavioral Therapy (CBT) on ICs' common psychological complaints, such as anxiety and depression, the objective was to conduct a meta-analysis on the effect of CBTs for adult ICs.

***Methods***: A literature search was conducted in order to identify all intervention studies on adult ICs that employed at least one therapeutic component defined as a CBT component.

***Results***: Literature searches revealed 36 unique records with sufficient data. These studies were subjected to meta-analyses using random effects models. A small, statistically significant effect of CBTs (Hedge's $g = 0.08$, $p = 0.014$) was revealed, which disappeared when randomized controlled trials were evaluated alone ($g = 0.04$, $p = 0.200$). A number of variables were explored as moderators. Only the percentage of female participants was positively associated with the effect size.

***Conclusions***: Based on the negligible effect of CBTs across outcomes, future studies should consider moving beyond traditional CBT methods as these do not appear efficacious. It is suggested that future interventions orient towards advances in the basic affective sciences and derived therapies in order to better understand and treat the emotional struggles experienced by ICs.



Received: 2 December 2015
Revised: 11 March 2016
Accepted: 25 March 2016

Ex

INTRODUCTION