

# **Statistics review**

Jonas Ranstam PhD

# Personal background

- Biostatistician and statistical epidemiologist
- BSc and PhD from Lund University (CStat from the RSS)
- 20 years university research (mainly observational studies)
- 10 years pharma and device industry research (mainly RCTs)
- Acta Orthopaedica since 1993
- Statistical Methods in Medical Research 2000 - 2004
- Osteoarthritis and Cartilage since 2008

# **Authors and editors have the same goals**

Advancement of scientific understanding and improvement in the treatment and prevention of disease.

Poor research methods, unnecessary research, redundant or duplicate publication, thinly sliced study results, selective reporting, and scientific fraud, as well as a general tendency to inflate the importance of the results, should all be resisted vigorously.

In particular, methodological review should be implemented much more widely.

Douglas G Altman. JAMA 2002;287:2765

**Table 1. Summary of Empirical Evidence of Prevalence of Methodological Problems in Published Reports of Randomized Trials\***

<b>Deficiency</b>	<b>Evidence</b>
Failing to specify eligibility criteria	25% of 364 reports in surgery journals
Not reporting an adequate method for generating random numbers	68% of 206 reports in obstetrics and gynecology journals; 52% of 80 reports in general medical journals
Not reporting the mechanism used to allocate interventions	89% of 196 reports in rheumatoid arthritis journals; 48% of 206 reports in obstetrics and gynecology journals; 44% of 80 reports in general medical journals
Failing to state whether blinding was used	51% of 506 reports in cystic fibrosis journals; 33% of 196 reports in rheumatoid arthritis journals; 38% of 68 reports in dermatology journals
Incorrect analysis of multiple observations	63% of 196 reports in rheumatoid arthritis journals
Inadequate information on harmful consequences of interventions	61% of 192 reports in 7 medical areas
Incorrect method of comparison of subgroups	58% of 50 reports in general journals

\*Data from Altman et al.<sup>2</sup>

Douglas G Altman. JAMA 2002;287:2765

# **Systematic reviews of orthopedic literature**

The ITT principle was recognized in 96 of 274 (35%) published randomized trials.

Herman A, Botser IB, Tenenbaum S, and Chechick A. Intention-to-treat analysis and accounting for missing data in orthopaedic randomized clinical trials. J Bone Joint Surg Am. 2009;91:2137-2143.

# Systematic reviews of orthopedic literature

A high proportion (42%) of clinical studies in high-impact-factor orthopedic journals involve the inappropriate use of multiple observations from single individuals

Bryant et al. How Many Patients? How Many Limbs?  
Analysis of Patients or Limbs in the Orthopaedic Literature.  
JBJS Am 2006;88:41-45.

# Systematic reviews of laboratory literature

A systematic review of 44 animal studies on fluid resuscitation shows ... that only two of the reviewed papers described how experimental units were allocated to treatment.

Roberts I, Kwan I, Evans P, Haig S. Does animal experimentation inform human healthcare? Observations from a systematic review of international animal experiments on fluid resuscitation. Br Med J 2002;324:474e6.

**Table 1.** Methodological Input in Relation to Study Design

Design	Methodologist, No./Total (%)		
	Biostatistician	Epidemiologist	Other
Randomized controlled trial	43/65 (66)	15/65 (23)	7/65 (11)
Systematic review	18/34 (53)	14/34 (41)	2/34 (6)
Observational	197/385 (51)	127/385 (33)	61/385 (16)
Economic	8/14 (57)	3/14 (21)	3/14 (21)
Other	7/16 (44)	3/16 (19)	6/16 (38)
<b>Total</b>	<b>273/514 (53)</b>	<b>162/514 (32)</b>	<b>79/514 (15)</b>

Altman et al. JAMA 2002;287:2817-2820

The ultimate interpretation and decision about the value of an article rests with the reader.

Gehlbach SH. Interpreting the Medical Literature: A Clinician's Guide. 3rd ed. New York, NY: McGraw-Hill; 1993.

## **The responsibilities of a statistical reviewer**

“To make sure that the authors spell out for the reader the limitations imposed upon the conclusions by the design of the study, the collection of data, and the analyses performed.”

Shor S. The responsibilities of a statistical reviewer. *Chest* 1972;61:486-487.

# **General statistical considerations**

regardless of whether an observational study or an experiment is described, and regardless of whether cells, animals or humans are studied.

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

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

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  $2 \times 2$  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.

# Study aim

## What is the aim of the study?

- To confirm a pre-specified hypothesis, or
- To generate new hypotheses

A confirmatory study protects the type-1 error rate, and this can only be achieved in an experiment (on cells, animals or patients).

## Do the authors' conclusion reflect their study aim?

# Statistical methods

## Evaluation of sampling/measurement uncertainty

- What methods were used?
- What assumption were made?
- How was the assumption fulfillment checked?
- Where any departures indicated?

# Results

## Observations

### Individual

- Dotplots, stem and leaf plots, scattergrams, etc.

### Aggregated

- Central tendency (mean, median, mode)
- Dispersion (SD, interquartile distance, range)
- Number of observations!!!!

# Results

## Interpretation

Uncertainty evaluation is based on

- 1) Effect/difference
- 2) Variability between independent observations
- 3) Number of observations!!!!

# **The number of observations is important**

27 hips from 21 patients or 130 pieces of cartilage from 5 patients indicate that not all observations are independent.

Many statistical techniques are based on an assumption of independent observations.

---

## Within- and Between-Subject Variation in Commonly Measured Anthropometric and Biochemical Variables

ADJI WIDJAJA,<sup>1</sup> RICHARD J. MORRIS,<sup>2</sup> JONATHAN C. LEVY,<sup>2\*</sup> KEITH N. FRAYN,<sup>3</sup>  
SUSAN E. MANLEY,<sup>2</sup> and ROBERT C. TURNER<sup>2</sup>

---

**Background:** The biological variation of some commonly assessed metabolic variables in healthy subjects has not been studied extensively. The aim of the study was to assess, in 12 healthy subjects (6 male and 6 female; mean (SD) age; 22.7 (1.5) years) following an overnight fast, the day-to-day variation of body fat (impedance method), triglycerides, nonesterified fatty acid (NEFAs), glycerol, 3-hydroxybutyrate (3-OHB), lactate, glucose, insulin (RIA), C-peptide, and glucagon on 12 consecutive days.

**Methods:** Between- and within-subject coefficients of variation ( $CV_G$  and  $CV_W$ ) were estimated using a random effects analysis of variance, and assay variation was subtracted to give the coefficient of within-subject biological variation ( $CV_I$ ). Individuality indices were calculated as  $CV_W/CV_G$ .

individual and group studies. The biological variation of some metabolites makes it difficult to characterize the status of healthy subjects with a single measurement.

© 1999 American Association for Clinical Chemistry

Physiological and metabolic studies commonly aim to describe the metabolic and biochemical status of individual subjects and to relate such variables to others, or to examine the effect of interventions. In many cases, it is assumed that the metabolic status of an individual may be represented by measurements taken on a single day, with allowance for confounding variables such as obesity where appropriate. Such studies often pay little attention to how such measurements fluctuate from day to day in healthy individuals. The importance of biological variation, however, is well recognized in the clinical chemistry

Within-subject variation

Between-subject variation

Analytic variability  
(measurement errors)

**Table 1. Between- and within-subject biological CVs and analytical CV of body composition, intermediary metabolites, and related hormones.**

Analyte	Overall mean or geometric mean	Biological CV		CV <sub>A</sub>
		CV <sub>G</sub>	CV <sub>I</sub>	
Body weight, kg	66.5	12%	0.9%	Negligible
Body fat, %	24.2	29%	10%	1.7% <sup>a</sup>
Glucose, mmol/L	4.94	7.5%	4.8%	2.4% <sup>b</sup>
Insulin, <sup>c</sup> pmol/L	52.2	26%	26%	6.6% <sup>b</sup>
C-peptide, <sup>c</sup> nmol/L	0.39	26%	24%	3.8% <sup>b</sup>
Glucagon, <sup>c</sup> ng/L	52.5	28%	19%	5.5% <sup>a</sup>
Triglycerides, <sup>c</sup> mmol/L	0.61	20%	21%	1.6% <sup>b</sup>
NEFAs, <sup>c</sup> μmol/L	376	32%	45%	4.1% <sup>b</sup>
Lactate, <sup>c</sup> mmol/L	0.88	29%	31%	1.6% <sup>a</sup>
Glycerol, <sup>c</sup> μmol/L	48.1	44%	36%	4.1% <sup>a</sup>
3-OHB, <sup>c</sup> μmol/L	42.5	41%	61%	6.1% <sup>a</sup>

<sup>a,b</sup>, CV <sup>a</sup> within- or <sup>b</sup> between-assay as appropriate.

<sup>c</sup> After log transformation.

# Results

Observations and interpretations should not be confused

- observed variability (SD) and
- estimation uncertainty (SEM)
  
- clinical significance (practical/clinical relevance)
- statistical significance (degree of uncertainty)

## Note

Only presenting p-values is not meaningful!!!

Results should be presented in terms of effect size, and their uncertainty should be indicated.

# Statistical vs. clinical significance

All statistically significant effects are not clinically important.  
(like a 6% reduction in body hair growth rate)

All clinically important effects are not statistically significant.  
(like a 20% increase in systolic blood pressure)

# Statistical vs. clinical significance

Clinical significance is statistically important for the

- i) sample size calculation,
- ii) performance of the statistical analysis, and
- iii) interpretation of the results.

# Uncertainty evaluation

## Hypothesis tests (p-values)

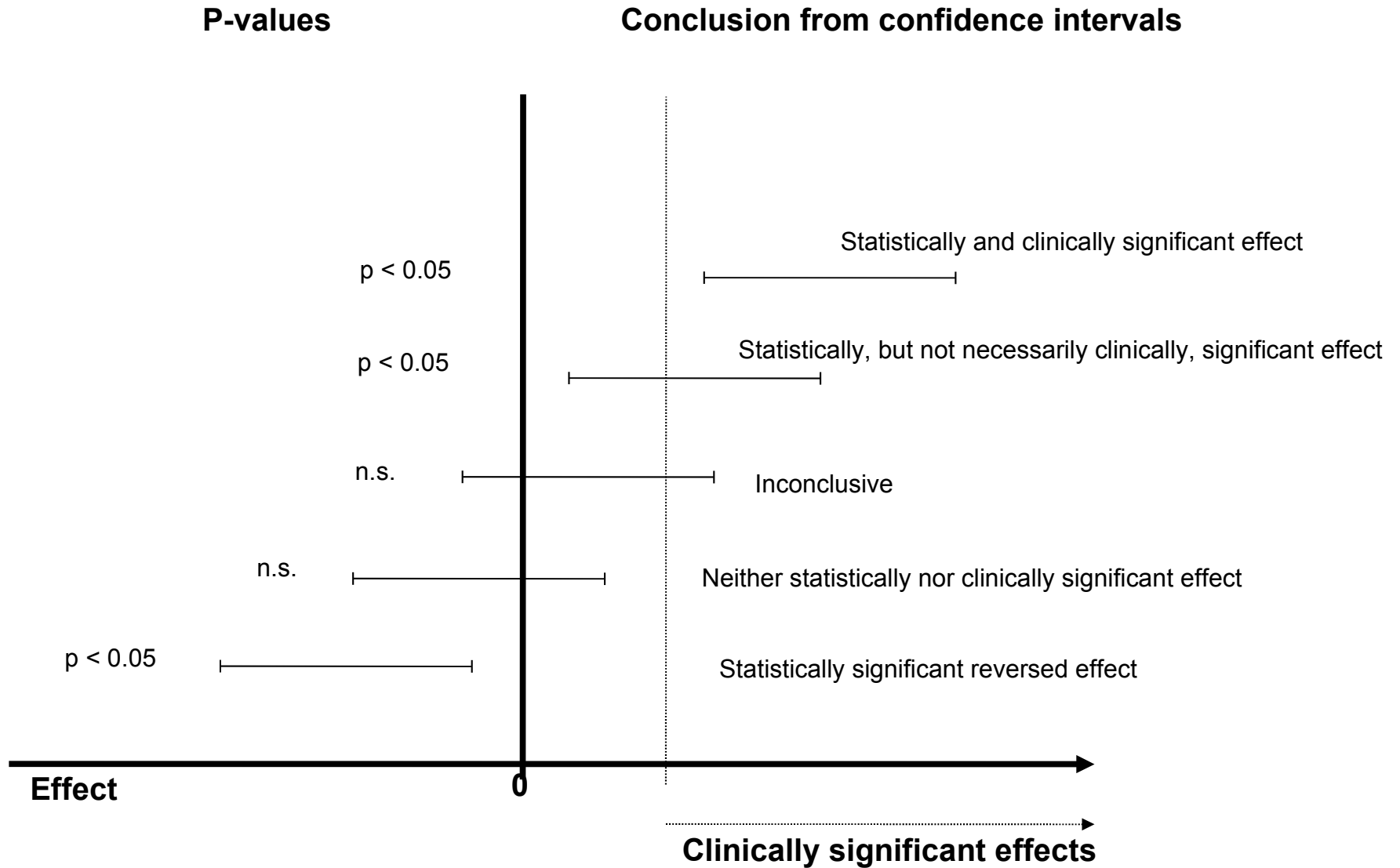
- Type-1 (only false positive) errors

## Interval estimation

- Confidence level, allows interpretation in terms of false positive as well as false negative errors.

(This is of course also important for the interpretation of differences with unknown clinical consequences)

# P-values vs. confidence intervals



# Results

## P-values

- should, unless  $p < 0.001$ , be presented as  $p = 0.023$
- not like  $p = 0.000$ ,  $p > 0.05$ ,  $p < 0.05$ , ns, or \*, \*\*, \*\*\*.

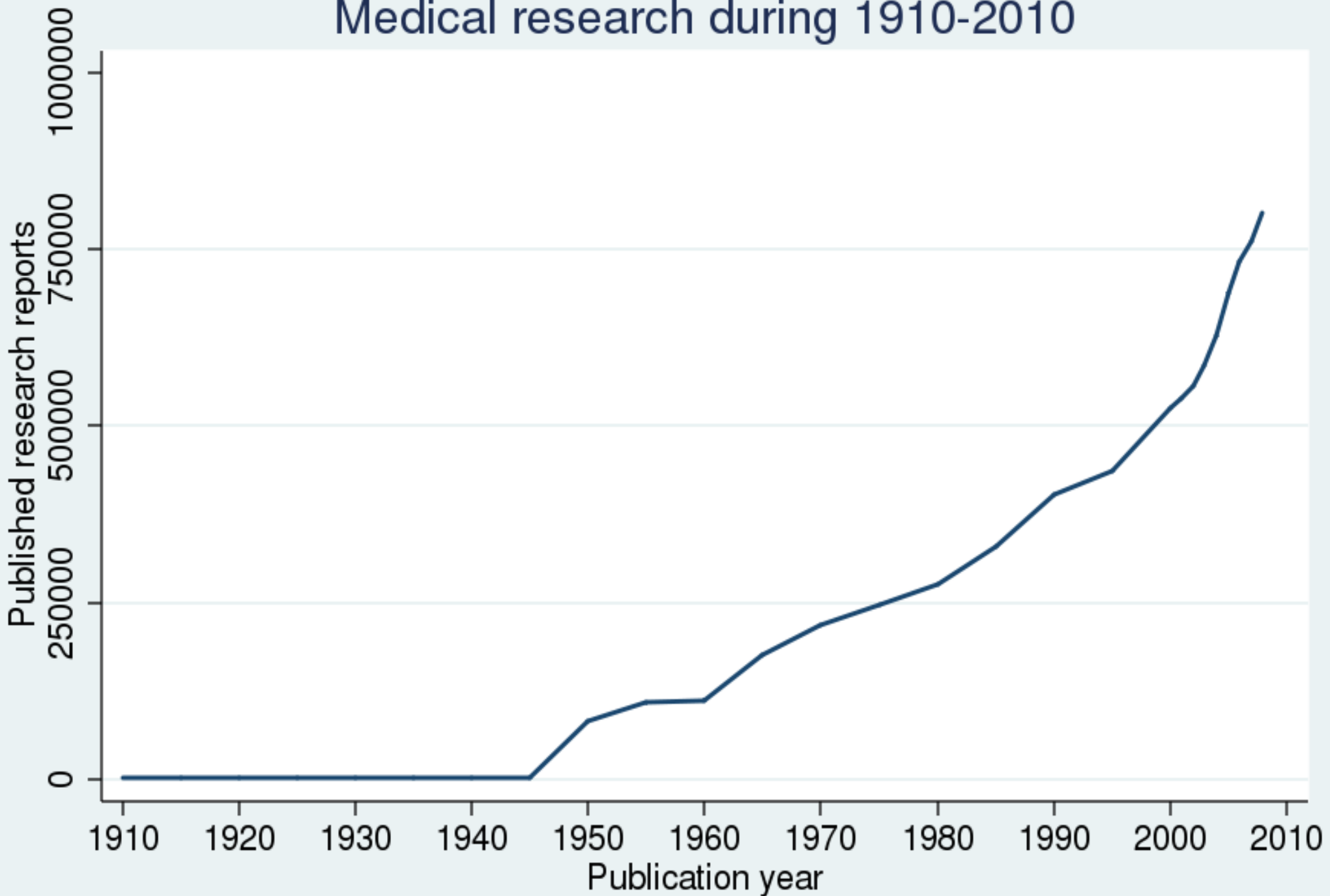
(Note that “ns” is the SI-unit for nanosecond)

## “Negative” results (statistical insignificance)

- “there was no difference”
- absence of evidence is not evidence of absence
- similarity may be presented using confidence intervals

**In what direction is medical  
research heading?**

# Medical research during 1910-2010



Source: US National Library of Medicine

# **Medical research as a modern science**

## **Randomised controlled trial (1948)**

Medical Research Council. Streptomycin in Tuberculosis Trials Committee. Streptomycin treatment of pulmonary tuberculosis. BMJ 1948;2:769-83.

## **Observational cohort study (1950)**

Doll R, Hill AB. Smoking and carcinoma of the lung. Preliminary report, BMJ 1950;2:739-748.

## **Case-control study (1954)**

Doll R, Hill AB. The mortality of doctors in relation to their smoking habits. BMJ 1954;228:1451-5

# Recent developments

## **A. ICH harmonized tripartite guidelines (1998)**

- E9 Statistical principles for clinical trials,
- Note for guidance, Points to consider, etc.

## **B. Public registration of study protocols (2005)**

- ClinicalTrials.gov, etc.
- WHO ICTRP

## **C. Reporting guidelines (1996 - 2010)**

- CONSORT (RTCs)
- PRISMA (Systematiska reviews)
- STROBE (Observationella studier)
- STARD (Diagnostiska studier)
- ARRIVE (Djurförsök)



## CONSORT 2010 checklist of information to include when reporting a randomised trial\*

Section/Topic	Item No	Checklist item	Reported on page No
<b>Title and abstract</b>			
	1a	Identification as a randomised trial in the title	_____
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	_____
<b>Introduction</b>			
Background and objectives	2a	Scientific background and explanation of rationale	_____
	2b	Specific objectives or hypotheses	_____
<b>Methods</b>			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	_____
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	_____
Participants	4a	Eligibility criteria for participants	_____
	4b	Settings and locations where the data were collected	_____
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	_____
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	_____
	6b	Any changes to trial outcomes after the trial commenced, with reasons	_____
Sample size	7a	How sample size was determined	_____
	7b	When applicable, explanation of any interim analyses and stopping guidelines	_____
Randomisation:			
Sequence generation	8a	Method used to generate the random allocation sequence	_____
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	_____
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	_____
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	_____
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those	_____

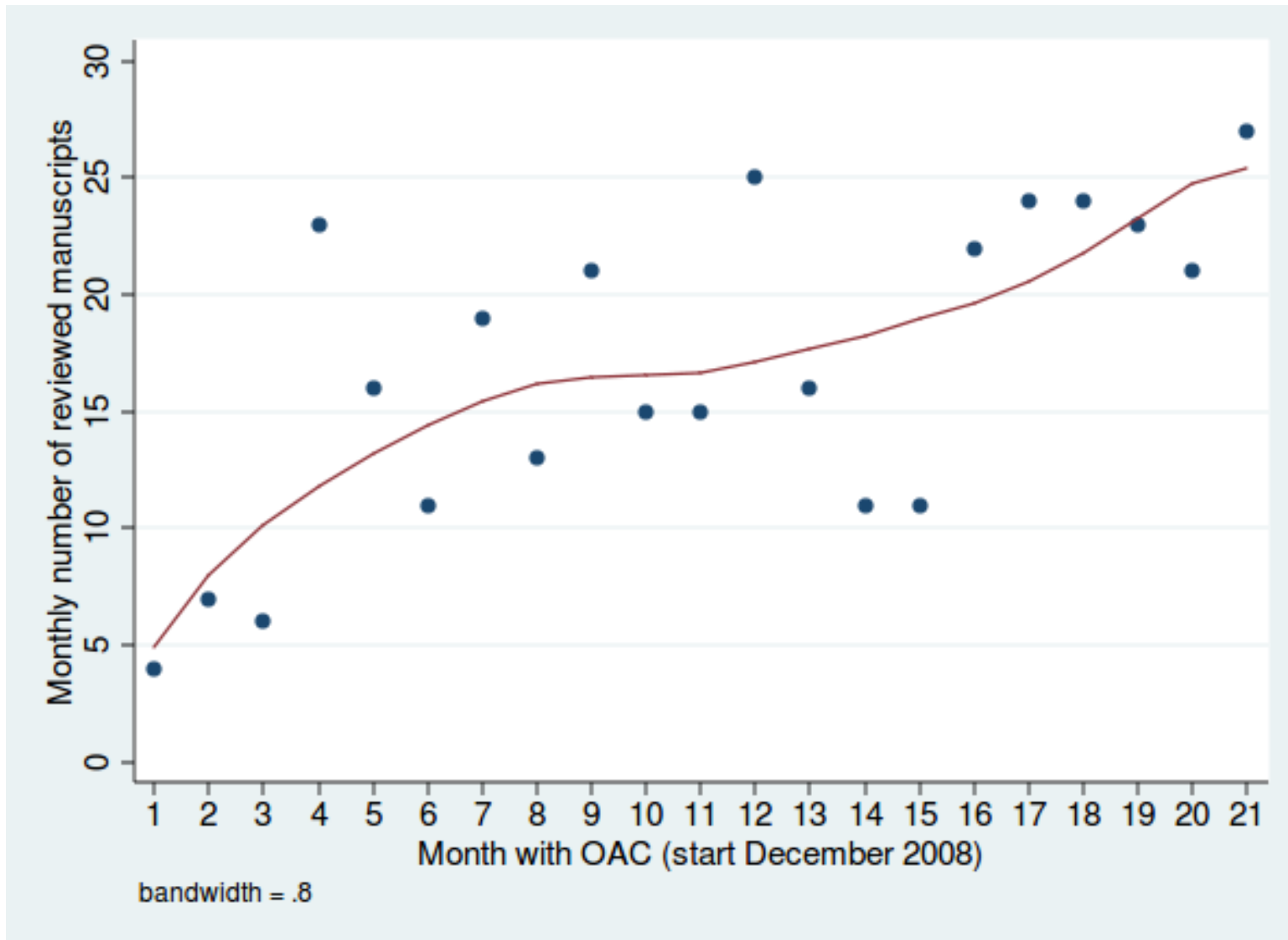
STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found
<b>Introduction</b>		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
<b>Methods</b>		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable,

**Table 2.** Animal Research: Reporting *In Vivo* experiments: The ARRIVE guidelines.

	ITEM	RECOMMENDATION
<b>TITLE</b>	1	Provide as accurate and concise a description of the content of the article as possible.
<b>ABSTRACT</b>	2	Provide an accurate summary of the background, research objectives (including details of the species or strain of animal used), key methods, principal findings, and conclusions of the study.
<b>INTRODUCTION</b>		
<b>Background</b>	3	a. Include sufficient scientific background (including relevant references to previous work) to understand the motivation and context for the study, and explain the experimental approach and rationale. b. Explain how and why the animal species and model being used can address the scientific objectives and, where appropriate, the study's relevance to human biology.
<b>Objectives</b>	4	Clearly describe the primary and any secondary objectives of the study, or specific hypotheses being tested.
<b>METHODS</b>		
<b>Ethical statement</b>	5	Indicate the nature of the ethical review permissions, relevant licences (e.g. Animal [Scientific Procedures] Act 1986), and national or institutional guidelines for the care and use of animals, that cover the research.
<b>Study design</b>	6	For each experiment, give brief details of the study design, including: a. The number of experimental and control groups. b. Any steps taken to minimise the effects of subjective bias when allocating animals to treatment (e.g., randomisation procedure) and when assessing results (e.g., if done, describe who was blinded and when). c. The experimental unit (e.g. a single animal, group, or cage of animals). A time-line diagram or flow chart can be useful to illustrate how complex study designs were carried out.
<b>Experimental procedures</b>	7	For each experiment and each experimental group, including controls, provide precise details of all procedures carried out. For example: a. How (e.g., drug formulation and dose, site and route of administration, anaesthesia and analgesia used [including monitoring], surgical procedure, method of euthanasia). Provide details of any specialist equipment used, including supplier(s). b. When (e.g., time of day). c. Where (e.g., home cage, laboratory, water maze). d. Why (e.g., rationale for choice of specific anaesthetic, route of administration, drug dose used).
<b>Experimental animals</b>	8	a. Provide details of the animals used, including species, strain, sex, developmental stage (e.g., mean or median age plus age range), and weight (e.g., mean or median weight plus weight range). b. Provide further relevant information such as the source of animals, international strain nomenclature, genetic modification status (e.g. knock-out or transgenic), genotype, health/immune status, drug- or test-naïve, previous procedures, etc.
<b>Housing and husbandry</b>	9	Provide details of: a. Housing (e.g., type of facility, e.g., specific pathogen free (SPF); type of cage or housing; bedding material; number of cage companions; tank shape and material etc. for fish). b. Husbandry conditions (e.g., breeding programme, light/dark cycle, temperature, quality of water etc. for fish, type of food, access to food and water, environmental enrichment). c. Welfare-related assessments and interventions that were carried out before, during, or after the experiment.

# Statistics review of OAC manuscript since December 2008



# Using resources more efficiently

Poorly written unclear manuscripts waste review resources:

- a) it takes too much time to understand what the authors have not described or explained
- b) clarifying what has actually been done takes one or more manuscript revisions, and these require further reviewing
- c) when the authors mistakes and misunderstandings have been clearly exposed the manuscript is rejected

# Using resources more efficiently

Author completed checklists will help authors write better manuscripts as well as facilitate the reviewing of them

**CONSORT** (RCTs)

**PRISMA** (Systematic reviews)

**STROBE** (Observational studies)

**STARD** (Studies on diagnostic accuracy)

**ARRIVE** (In vitro experiments)

**Thank you for your attention!**