Measures of Fidelity in Interpretation and Communication

Isabelle Boutron

Centre of Research in Epidemiology and Statistics

Paris Cité University

Université Paris Cité

AP-HP. Centre Université de Paris

Inserm
La science pour la santé
From science to health

Reporting, interpretation and communication of research results

- Ethical principles for medical research involving human subjects (Declaration of Helsinki)
  - Researchers have a duty to **make publicly available** the results of their research on human subjects and are accountable for the **completeness and accuracy** of their reports.
- Inaccurate reporting, interpretation and communication can have dramatic effect
  - COVID-19 pandemic – Hydroxychloroquine

Gould, Norris Cochrane Database of Systematic Reviews 2021

Distorted reporting, interpretation and communication

Selective reporting

- **Selective publication**
  - Studies with statistically significant results
- **Selective reporting of outcome and analyses**
  - ‘P-hacking’

Turner et al. NEJM 2008
Chan et al. JAMA 2004

Spin

- A **specific reporting** that **fails to faithfully reflect the findings** and that could affect the impression that the results produce in readers
- Empirical evidence shows that spin can **impact** readers’ interpretation

Boutron , Ravaud. PNAS 2018
Boutron et al. J Clinical Oncology 2014
Boutron et al. BMC Med 2019

Measurement of selective reporting of outcomes and analyses

- **Measurement**
  - Comparison of what was planned (protocol/registries) to what was reported
- **Prevalence of selective outcome**
  - Comparison of registry record and published reports in 206 COVID-19 RCTs

  -> inconsistency in primary outcomes in **97 (47%) trials**

Outcome Switch	N=97
Added primary outcome(s)	8%
Removed primary outcome(s)	19%
Added and removed primary outcome(s)	17%
Changed primary outcome(s)	36%
Time frame or metric different	21%

Kapp et al., BMC in Revision

Measurement of spin

JAMA

Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes

Isabelle Boutron, MD, PhD
Susan Dutton, MSc
Philippe Ravaud, MD, PhD
Douglas G. Altman, DSc

META-RESEARCH ARTICLE

PLOS BIOLOGY

'Spin' in published biomedical literature: A methodological systematic review

Kellia Chiu, Quinn Grundy, Lisa Bero*

Charles Perkins C

* lisa.bero@sydn

Lazarus *et al. BMC Medical Research Methodology* (2015) 15:85
DOI 10.1186/s12874-015-0079-x

BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Classification and prevalence of spin in abstracts of non-randomized studies evaluating an intervention

Clément Lazarus[1,2,3], Romana Haneef[1,2], Philippe Ravaud[1,2,3,4,5] and Isabelle Boutron[1,2,3,4*]

Radiology

Eleanor A. Ochodo, MBChB, MIH
Margriet C. de Haan, MD
Johannes B. Reitsma, MD, PhD
Lotty Hooft, PhD
Patrick M. Bossuyt, PhD
Mariska M. G. Leeflang, PhD

Overinterpretation and Misreporting of Diagnostic Accuracy Studies: Evidence of "Spin"[1]

Clinical Chemistry 66:7
915-924 (2020)

Overinterpretation of Research Findings: Evaluation of "Spin" in Systematic Reviews of Diagnostic Accuracy Studies in High-Impact Factor Journals

Trevor A. McGrath,[a] Joshua C. Bowdridge,[b] Ross Prager,[b] Robert A. Frank,[a] Lee Treanor,[c] Ana Dehmoobad Sharifabadi,[c] Jean-Paul Salameh,[d] Mariska Leeflang,[e] Daniël A. Korevaar,[f] Patrick M. Bossuyt[e], and Matthew D. F. McInnes[a,d,*]

ELSEVIER

Check for updates

Journal of Clinical Epidemiology 116 (2019) 9–17

Journal of Clinical Epidemiology

REVIEW

A systematic review finds that spin or interpretation bias is abundant in evaluations of ovarian cancer biomarkers

Mona Ghannad[a,b,*], Maria Olsen[a,b], Isabelle Boutron[b], Patrick M. Bossuyt[a]

[a]Amsterdam UMC, Department of Clinical Epidemiology, Biostatistics and E
Institute, Meibergdreef 9, 1105 A

Open access Original research

BMJ Open

Cross-sectional study of preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and spin in interpretation

Lisa Bero ,[1] Rosa Lawrence,[2] Louis Leslie,[2] Kellia Chiu,[3] Sally McDonald,[3] Matthew J Page ,[4] Quinn Grundy ,[5] Lisa Parker ,[6] Stephanie Boughton,[7] Jamie J Kirkham ,[8] Robin Featherstone[7]

ELSEVIER

CrossMark

Journal of Clinical Epidemiology 75 (2016) 56–65

Journal of Clinical Epidemiology

A new classification of spin in systematic reviews and meta-analyses was developed and ranked according to the severity

Amélie Yavchitz[a,b,c,*], Philippe Ravaud[a,b,c,d], Douglas G. Altman[e], David Moher[f,g], Asbjørn Hrobjartsson[h,i], Toby Lasserson[j], Isabelle Boutron[a,b,c]

[a]Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153), Inserm/Université Paris Descartes, 1 place du Parvis Notre Dame, Paris 75004, France
[b]Centre d'Épidémiologie Clinique, AP-HP (Assistance Publique des Hôpitaux de Paris), Hôpital Hôtel Dieu, 1 place du Parvis Notre Dame, Paris, France
[c]French Cochrane Center, Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, 1 place du Parvis Notre Dame, Paris 75004, France

Measurement of spin

Methods for measuring spin

- Systematic assessment of published reports (10-374 reports assessed)
- Prespecified standardized data collection instrument developed by authors or inductive methods
- Subjectivity in the assessment
- Multiple independent assessors with consensus

Chiu et al. Plos Biology 2017

Spin practices measured

Misreporting reporting of methods

- Changed objective and hypothesis
- Methods beautification
- No identification of pre-specified and post-hoc analysis
- Failure to acknowledge protocol deviation

Misreporting reporting of results

- Selective focus on outcomes favoring the study hypothesis
- Ignoring or understating results contradicting the initial hypothesis (e.g., adverse events)
- Figures/images misrepresenting the data
- Use of linguistic spin

Misinterpretation

- Misleading interpretation (ignoring regression to the mean, confounding, small study effect)
- Misinterpretation of p-value as a measure of effect, lack of statistical significance as demonstrating equivalence or safety
- Ignoring limitations

Inaccurate extrapolation

- Extrapolation to larger population, different setting, different outcome, set of interventions,
- Providing recommendations not supported by the data

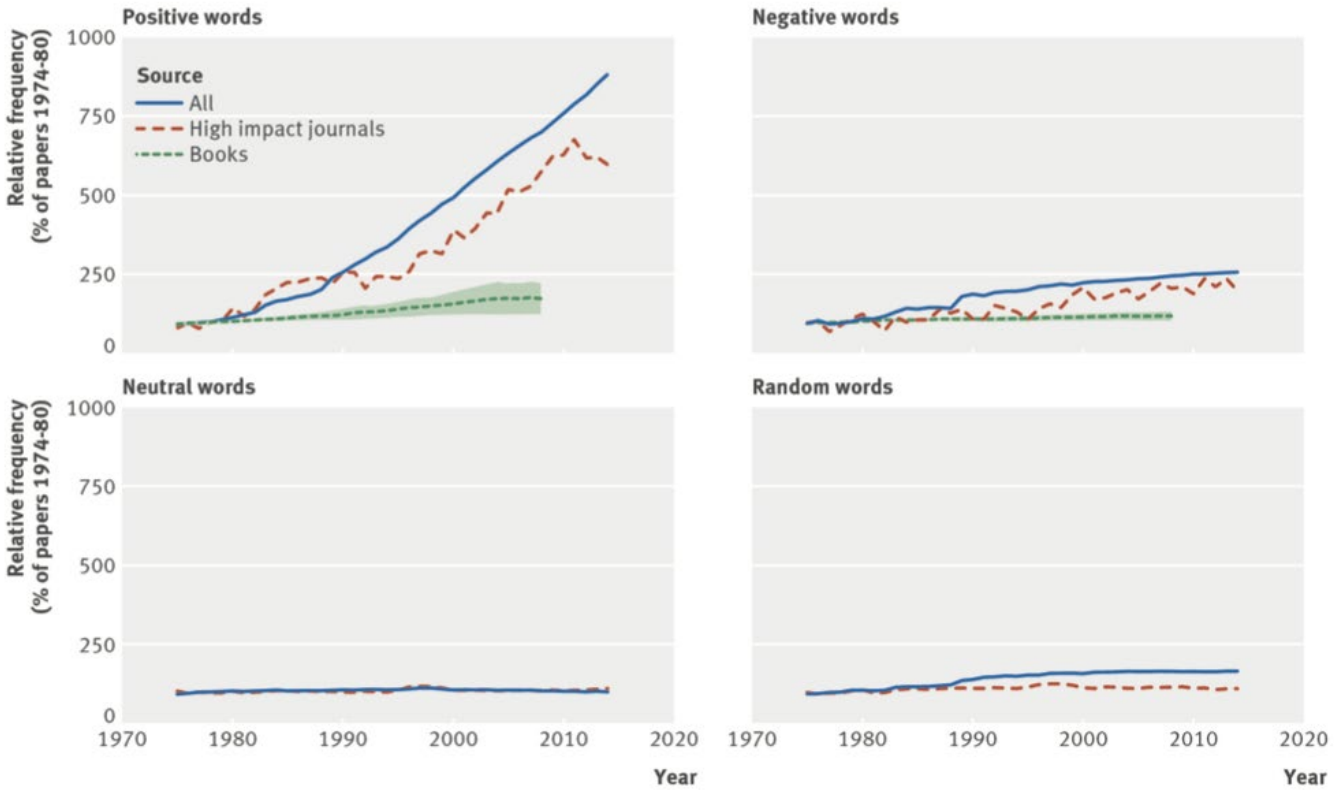
Prevalence of spin

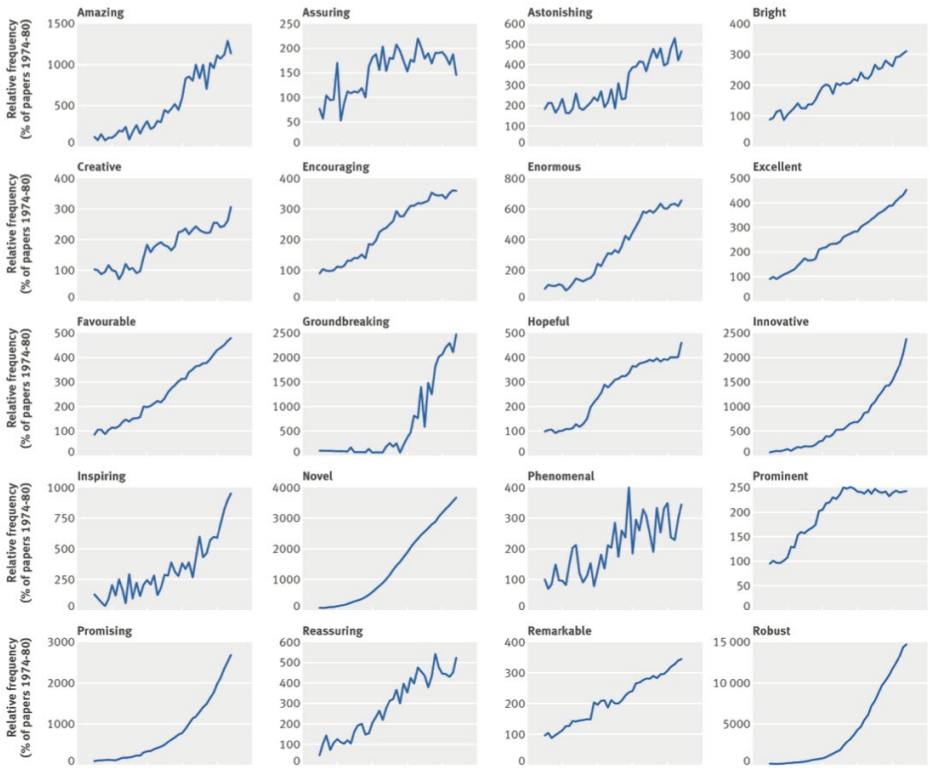
	Spin in abstract Median% (Min-Max%) (*n* measures)	Spin in full text Median% (Min-Max%) (*n* measures)
Trials	56.8 (9.7–83.6) (*n*=13)	56.5 (18.8–100) (*n*=16)
Observational studies	30.7 (23.9–38.6) (*n*=3)	85.6 (85.6–85.6) (*n* = 1)
Diagnostic accuracy studies	-	43.7 (31.0–56.5) (*n* = 2)
Systematic reviews/meta-analyses	-	26.3 (24.2–28.4) (*n* = 2)

Chiu et al. Plos Biology 2017

Lexicographic analysis – linguistic spin

- To investigate whether language used in science abstracts can skew towards the use of strikingly positive and negative words over time.





Vinkers et al. BMJ 2015

DeSpin: a prototype system for detecting spin and selective reporting in biomedical publications

- Natural Language Processing (NLP) system for detecting spin in biomedical articles reporting randomized controlled trials (RCTs).
  - Identification sub-types of spin separately
  - E.g., Focus on the switching of the primary outcome.
    - Identification of primary outcome in registry and article
    - Assessment of semantic similarities (abstract, text, registry)
    - Discourse prominence of the reported primary outcome

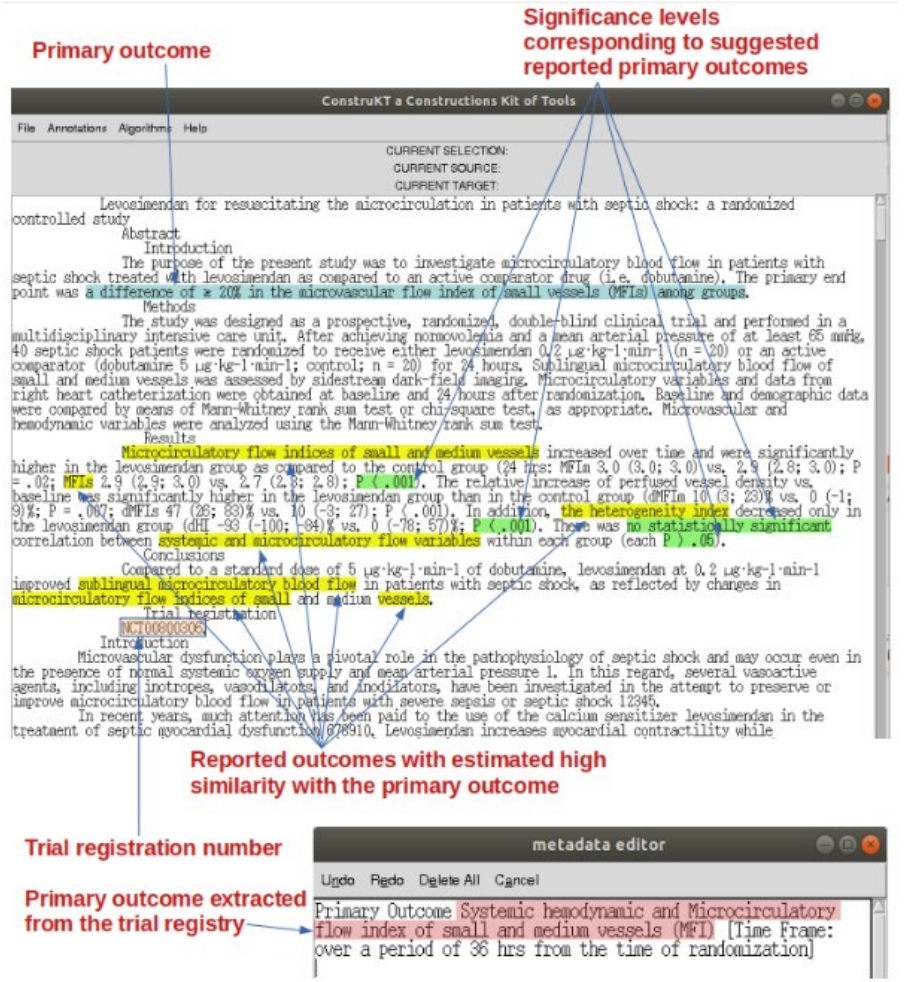
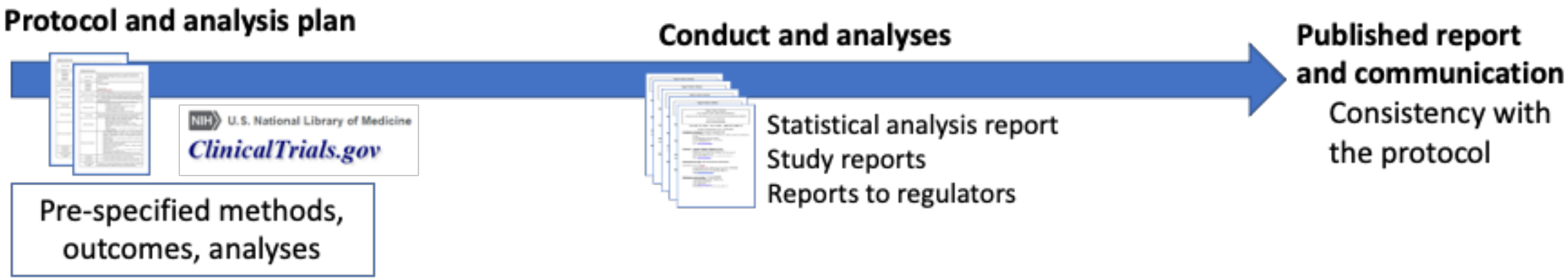


Figure 1: Example of a processed text

Koroleva A et al., *Proceedings of the BioNLP 2020 workshop*,

Barriers to measurement

- Lack of access to protocols and all research documentations
- Incomplete reporting in available resources (e.g., registries)
- Complexity and subjectivity of measurement (spin)

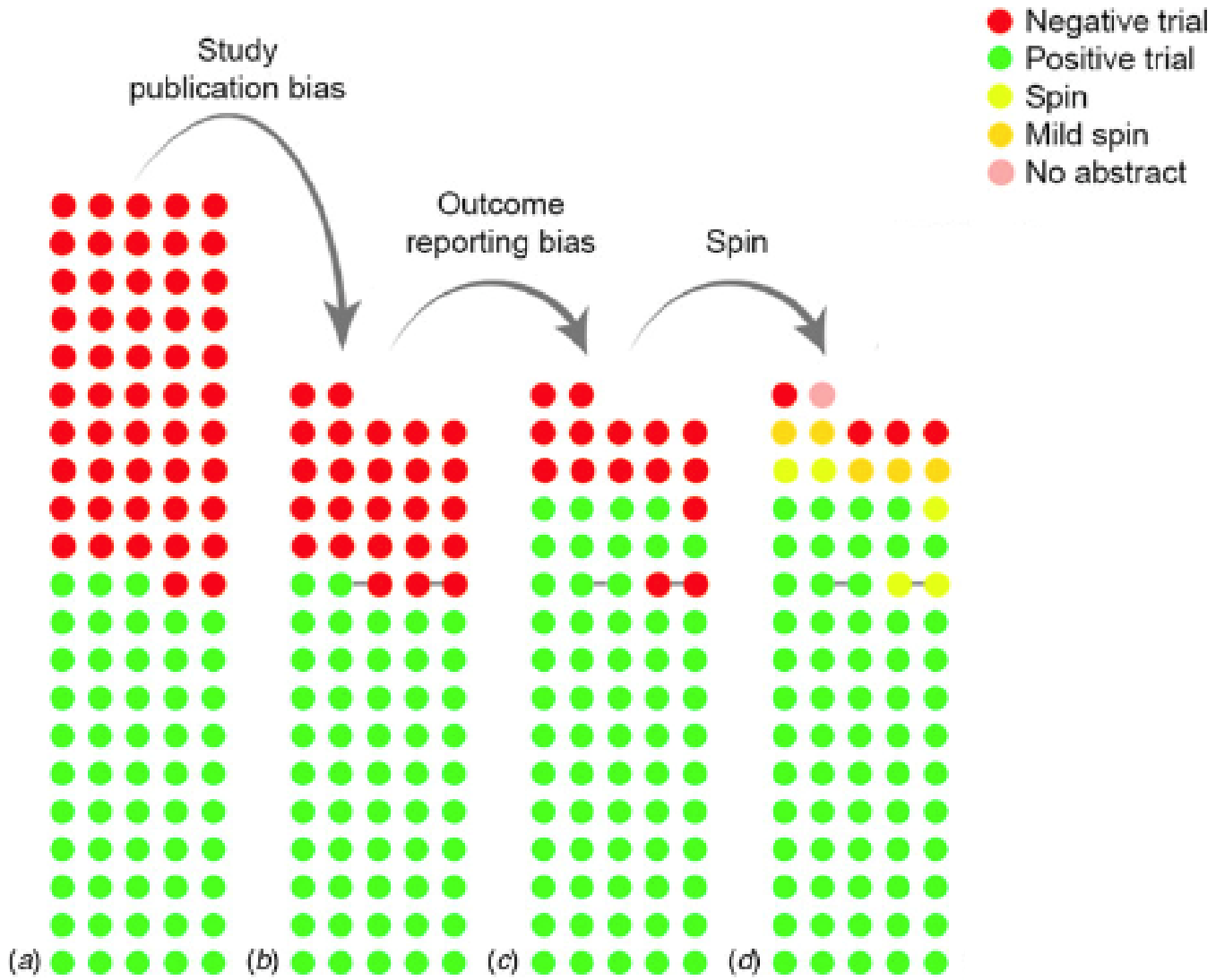


Conclusions

- Inaccurate communication of research results can have dramatic consequences.
- The current system is not sufficiently efficient to detect and correct inaccurate reporting and interpretation.
- Tools to measure and detect selective reporting and spin have been developed.
- We need to move toward tools allowing large scale detection

Example

Cohort of 105 depression trials approved by the FDA

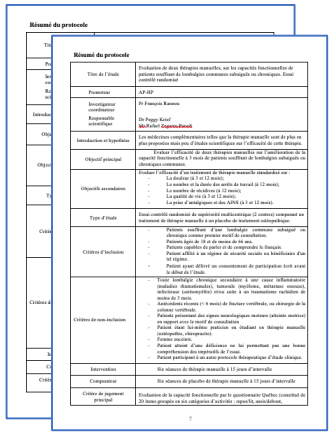


de Vries et al Psychological Medicine 2018 – figure adapted
Turner et al. NEJM 2008

Interpretation and communication of research results

Planning

Protocol, and analysis plan


Résumé du protocole

Titre de l'étude	Evaluation de deux thérapies manuelles, sur les capacités fonctionnelles de patients souffrant de lombalgies communes subaiguës ou chroniques. Essai contrôlé randomisé
Promoteur	AP-HP
Investigateur coordinateur Responsable scientifique	Pr François Rannou Dr Peggy Krief Mr Robert Eugene Ewodi
Introduction et hypothèse	Les médecines complémentaires telles que la thérapie manuelle sont de plus en plus proposées mais peu d'études scientifiques sur l'efficacité de cette thérapie.
Objectif principal	Evaluer l'efficacité de deux thérapies manuelles sur l'amélioration de la capacité fonctionnelle à 3 mois de patients souffrant de lombalgies subaiguës ou chroniques communes.
Objectifs secondaires	Evaluer l'efficacité d'un traitement de thérapie manuelle standardisé sur : - La douleur (à 3 et 12 mois); - Le nombre et la durée des arrêts de travail (à 12 mois); - Le nombre de récidives (à 12 mois); - La qualité de vie (à 3 et 12 mois); - Le prise d'antalgiques et des AINS (à 3 et 12 mois).
Type d'étude	Essai contrôlé randomisé de supériorité multicentrique (2 centres) comparant un traitement de thérapie manuelle à un placebo de traitement orthopédique.
Critères d'inclusion	- Patients souffrant d'une lombalgie commune subaiguë ou chronique comme premier motif de consultation. - Patients âgés de 18 et de moins de 66 ans. - Patients capables de parler et de comprendre le français. - Patient affilié à un régime de sécurité sociale ou bénéficiaire d'un tel régime. - Patient ayant délivré un consentement de participation écrit avant toute procédure.
Critères de non-inclusion	- Toute lombalgie chronique secondaire à une cause inflammatoire (maladies rhumatismales), tumorale (myélome, métastase osseuse), infectieuse (ostéomyélite) et/ou suite à un traumatisme rachidien de moins de 3 mois. - Antécédents récents (< 6 mois) de fracture vertébrale, ou chirurgie de la colonne vertébrale. - Patient présentant des signes neurologiques moteurs (atteinte motrice) à examiner avec le motif de consultation. - Patient ayant bénéficié précédemment ou réalisant en thérapie manuelle (ostéopathie, chiropractie). - Femme enceinte. - Patient atteint d'une déficience ne lui permettant pas une bonne compréhension des impératifs de l'essai. - Patient participant à un autre protocole thérapeutique d'étude clinique.
Intervention	Six séances de thérapie manuelle à 15 jours d'intervalle
Comparateur	Six séances de placebo de thérapie manuelle à 15 jours d'intervalle
Critère de jugement principal	Evaluation de la capacité fonctionnelle par le questionnaire Québec (constitué de 20 items groupés en six catégories d'activités : rapport/b, sociabilité,

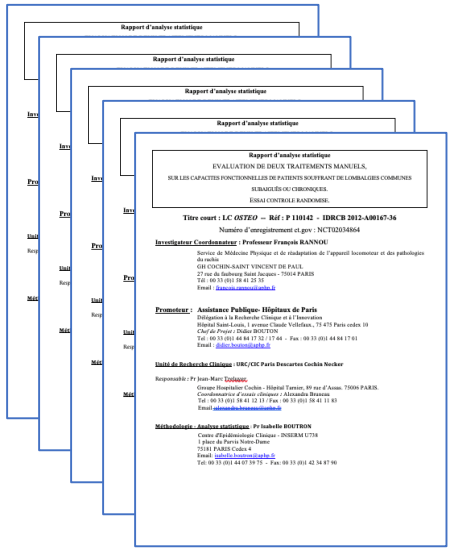
NIH U.S. National Library of Medicine

ClinicalTrials.gov

Pre-specified hypothesis, methods, outcomes, analyses

Conduct and analyses

- Statistical analysis report
- Study reports
- Reports to regulators


Rapport d'analyse statistique

EVALUATION DE DEUX TRAITEMENTS MANUELS,
SUR LES CAPACITES FONCTIONNELLES DE PATIENTS SOUFFRANT DE LOMBALGIES COMMUNES
SUBAIGUËS OU CHRONIQUES.
ESSAI CONTROLE RANDOMISE.

Titre court : *LC OSTEO -- Réf : P 110142 - IDRCB 2012-A00167-36*

Numéro d'enregistrement et.gov : NCT02034864

Investigateur Coordonnateur : Professeur François RANNOU

Service de Médecine Physique et de réadaptation de l'appareil locomoteur et des pathologies du rachis
GH COCHIN-SAINT VINCENT DE PAUL
27 rue du faubourg Saint Jacques - 75014 PARIS
Tél : 00 33 (0)1 58 41 25 35
Email : francois.rannou@aphp.fr

Promoteur : Assistance Publique- Hôpitaux de Paris
Délégation à la Recherche Clinique et à l'Innovation
Hôpital Saint-Louis, 1 avenue Claude Vellefaux, 75 475 Paris cedex 10
Chef de Projet : Didier BOUTRON
Tel : 00 33 (0)1 44 84 17 32 / 17 44 - Fax : 00 33 (0)1 44 84 17 01
Email : didier.boutron@aphp.fr

Unité de Recherche Clinique : URC/CIC Paris Descartes Cochin Necker

Responsable : Pr Jean-Marc Tredaniel
Groupe Hospitalier Cochin - Hôtel-Dieu - Broca - La Roche Guyon
Coordonnatrice d'essais cliniques : Alexandra Brasseur
Tel : 00 33 (0)1 58 41 12 13 / Fax : 00 33 (0)1 58 41 11 83
Email : alexandra.brasseur@cch.ap.fr

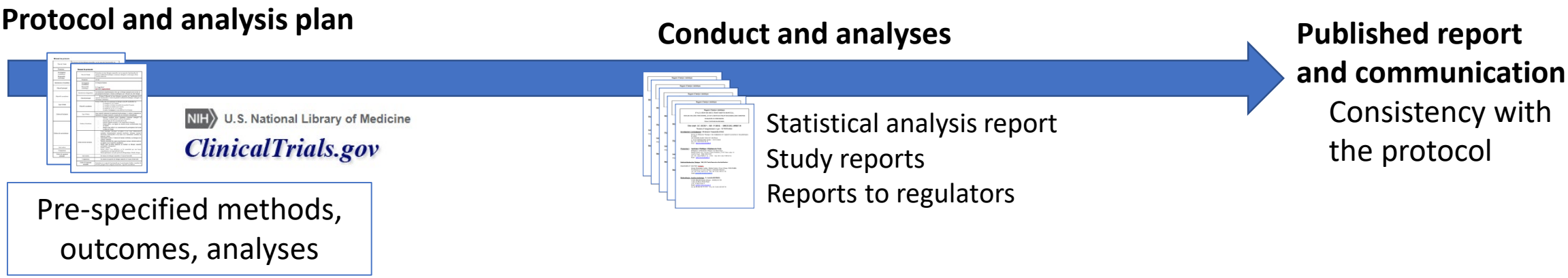
Méthodologie - Analyse statistique : Pr Isabelle BOUTRON
Centre d'Epidémiologie Clinique - INSERM U738
1 place du Parvis Notre-Dame
75181 PARIS Cedex 4
Email : isabelle.boutron@aphp.fr
Tel : 00 33 (0)1 44 07 39 75 – Fax : 00 33 (0)1 42 34 87 96

Accurate reporting and interpretation

- Report pre-specified methods
- Report all deviations from the protocol
- Report results of pre-specified analyses
- Focus on results of pre-specified primary analysis
- Report appropriate inference
- Be caution to avoid misinterpretation or inadequate extrapolations
- Highlight limitations

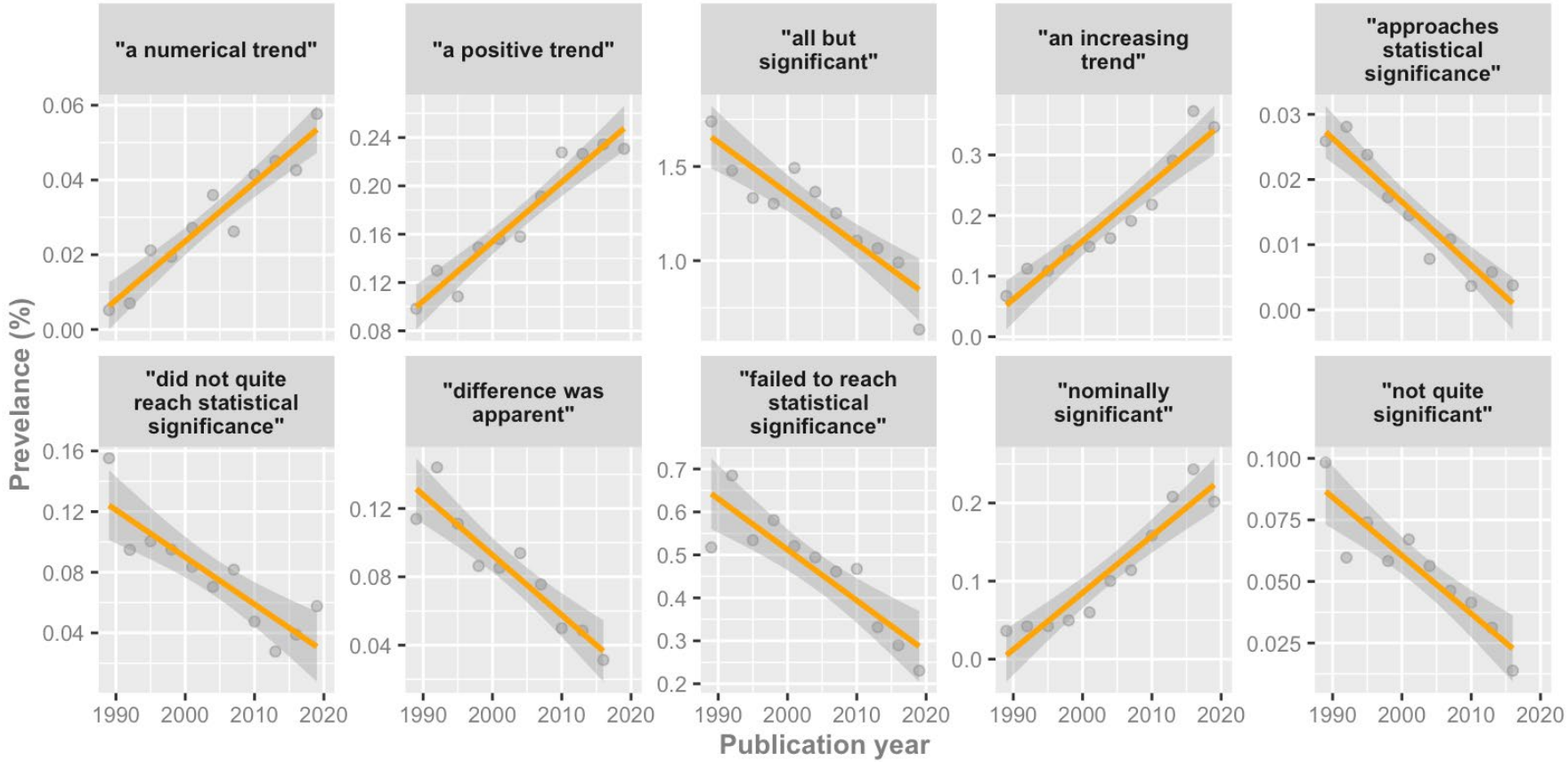
Reporting, interpretation and communication of research results

- Ethical principles for medical research involving human subjects (Declaration of Helsinki)
  - Researchers have a duty to **make publicly available** the results of their research on human subjects and are accountable for the **completeness and accuracy** of their reports.
- Inaccurate reporting, interpretation and communication can have dramatic effect
  - COVID-19 pandemic – Hydroxychloroquine



Trends in phrases used to discuss results that do not reach statistical significance – linguistic spin

- 505 predefined phrases denoting results that approach but do not cross the line of formal statistical significance
- 567,758 RCTs recorded in PubMed between 1990 and 2020



Otte Plos BIology 2022