



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

LONG-TERM IMPACT OF AN EDITORIAL INTERVENTION TO IMPROVE PAPER TRANSPARENCY AND REPRODUCIBILITY

Marta Vilaró Pacheco

Department of Statistics and Operations Research

Doctoral Thesis, 2021

Thesis directors:

Erik Cobo Valeri, PhD

José Antonio González Alastrué, PhD

Department of Statistics and Operations Research

Universitat Politècnica de Catalunya, Spain

“To maximize the benefit to society, you need to not just do research but do it well.”

Professor Doug Altman

Medical researcher and statistics game-changer

([EQUATOR network](#))

ACKNOWLEDGEMENTS

Primer de tot vull agrair aquesta tesis als que considero els meus mentors, l'Erik i el José Antonio. Crec que ells han cregut en mi més que jo en mi mateixa en molts moments. Amb els seu suport i els seus consells he aconseguit tirar endavant aquest treball. M'han ensenyat tant d'estadística i rigorositat científica com de gestió del temps. També els agraeixo tots els cafès que hem compartit, allà he après sobre la vida.

També vull agrair als qui m'han acompanyat des del DEIO. Als meus companys Jordi, Marta, David, Ceci i Yovi. Amb ells he compartit seminaris i algun congrés, i moments molt enriquidors, m'hagués agradat estar més present. No em puc descuidar la Sonia, gràcies per tots els dubtes de gestió que m'has resolt.

Gràcies al GRBIO i en especial als professors Lupe i Klaus amb els que he compartit moments de departament i algun congrés.

També vull agrair al Matt el seu esforç en la revisió de l'anglès de la tesis.

Als meus pares i al meu germà que sempre m'han impulsat a aconseguir el que em proposava.

I per acabar vull agrair la tesis a les persones més especials i que m'omplen la vida. Al meu encant, que m'ha acompanyat en tot el transcurs de la tesis des de la il·lusió del principi a l'alegria del final, passant per moments durs, gràcies Vane. I als dos petitons que han arribat en mig de la il·lusió i l'alegria, gràcies Pau i Alba per ser pura energia i alegria. M'heu ensenyat què és l'amor infinit.

ABSTRACT

This PhD thesis is concerned with evaluating the long-term effects of interventions in the editorial peer review process of a biomedical journal. The growing need to increase the value of research and avoid waste motivates this work. We focus on evaluating the long-term impact on the number of citations (NC) of articles introduced in trials, as this will allow us to evaluate the effects of adding either a methodological expert (statistician) or reporting guidelines (RG) during the peer review process.

In 2004 and 2009, two randomized trials were published in *Medicina Clínica* (Elsevier, Barcelona), in which these interventions were added during the editorial process. They showed a slightly positive effect on paper quality, as assessed by the Manuscript Quality Assessment Instrument (MQAI) of Goodman et al. In this work, we explore the effect on NC by collecting from Web of Science (WoS) the NC of each article that had been randomized in the two previous studies.

This thesis presents different ways of analyzing count data involving time-to-event, first from a perspective of count data models and then from a perspective of recurrent events methods.

Results show that including a methodological reviewer (for example, a senior statistician) who is dedicated to looking for missing RG items increases the NC by 40% (95% CI: 1% to 94%) when considering that citations are independent within articles, like count models. When considering the within-article correlation of citations using the frailty gamma model, we find that including a methodological reviewer during the peer-review process increases the possibility of receiving a citation by 41% (95% CI: 0% to 102%). Independently of the method used to analyze NC, and with different assumptions, the consistency of those results gives robustness to the findings.

Therefore, as measured by the NC, with this work we show that randomized studies of using interventions in the peer review process to improve scientific impact are feasible. Our proof of concept study opens the door for the development of confirmatory trials.

Keywords: Reporting guidelines, peer review, reproducibility, number of citations, count models, recurrent events models, research on research

RESUMEN

El interés de esta tesis recae en la evaluación del efecto a largo plazo de intervenciones propuestas durante el proceso editorial de revisión por pares de una revista biomédica. La creciente necesidad para incrementar el valor de la investigación evitando gastos motiva este trabajo. Nos hemos centrado en evaluar el impacto a largo plazo mediante el número de citas (NC) de unos artículos que forman parte de dos ensayos clínicos, lo que nos permite evaluar el efecto de añadir tanto un experto en metodología (estadístico) como las guías de publicación (RG) durante el proceso editorial de revisión por pares.

En 2004 y 2009 se publicaron dos ensayos clínicos en la revista biomédica *Medicina Clínica* (Elsevier, Barcelona), dónde se implementaron estas intervenciones durante el proceso editorial. Sus autores encontraron un ligero efecto positivo en la calidad del manuscrito, evaluada mediante el instrumento de evaluación de la calidad de los manuscritos (MQAI) de Goodman. En este trabajo, exploramos el efecto sobre el NC recogiendo de la *Web of Science* (WoS) el NC de cada artículo aleatorizado en los dos estudios previos.

Esta tesis presenta diferentes formas de analizar datos de contaje dónde interviene el tiempo hasta el evento, primero desde una perspectiva de modelos de datos de contaje y después desde una perspectiva de modelos de eventos recurrentes.

Los resultados muestran que incluir un revisor metodológico (por ejemplo, un estadístico senior) que se dedique a buscar ítems ausentes de las RG incrementa el NC un 40% (95% IC: 1% a 94%) cuando consideramos que las citas son independientes dentro de los artículos, como en los modelos de contaje. Cuando consideramos la correlación intra-artículo usando el modelo gamma de fragilidad, encontramos que incluir un revisor metodológico durante el proceso editorial de revisión por pares incrementa la posibilidad de recibir una cita un 41% (95% CI: 0% a 102%). Independientemente de la metodología utilizada para analizar el NC, y con diferentes asunciones, la consistencia de estos resultados da robustez a los hallazgos.

Con este trabajo mostramos que son factibles los ensayos clínicos que implementan intervenciones en el proceso editorial de revisión por pares para mejorar el impacto científico. Nuestro estudio abre la puerta al desarrollo de estudios confirmatorios en este ámbito.

Palabras clave: guías de publicación, revisión editorial por pares, reproducibilidad, número de citas, modelos de contaje, modelos de eventos recurrentes, investigación en investigación

CONTENTS

LIST OF FIGURES	iii
LIST OF TABLES	v
ABBREVIATIONS	vi
CHAPTER 1: INTRODUCTION	8
1.1. GENERAL INFORMATION	8
1.2. DISSERTATION STRUCTURE	11
1.3. STATE OF EXISTING RESEARCH	13
1.3.1. EDITORIAL PEER REVIEW PROCESS	13
1.3.2. REPORTING GUIDELINES	17
1.3.3. RESEARCH IMPACT	20
1.3.5. INTERVENTIONS TO IMPROVE MANUSCRIPTS THROUGHOUT EDITORIAL PEER REVIEW IN BIOMEDICAL JOURNALS	23
1.4. MAIN AND SECONDARY OBJECTIVES	29
CHAPTER 2: METHODOLOGY	30
2.1. DATA COLLECTION	30
2.2. COUNT DATA MODELS	34
2.2.1. NON-PARAMETRIC EFFECT ESTIMATION	34
2.2.2. POISSON REGRESSION MODEL	35
2.2.3. QUASI-POISSON AND NEGATIVE BINOMIAL REGRESSION MODEL	38
2.2.4. ZERO-INFLATED POISSON AND ZERO-INFLATED NEGATIVE BINOMIAL MODELS	39
2.2.5. COMPARING MODELS	41
2.3. JOINT EFFECT ESTIMATION	43
2.4. RECURRENT EVENTS METHODS	44

2.4.1. MEAN CUMULATIVE FUNCTION	45
2.4.2. FRAILTY MODELS	46
2.5. REPRODUCIBILITY (Shiny application)	51
CHAPTER 3. RESULTS	53
3.1. DESCRIPTIVE RESULTS.....	53
3.2. EFFECT OF A STATISTICAL REVIEWER USING REPORTING GUIDELINES	57
3.3. SEPARATE EFFECTS OF STATISTICAL REVIEWER AND REPORTING GUIDELINES	68
CHAPTER 4: DISCUSSION.....	69
4.1. MAIN FINDINGS	69
4.2. LIMITATIONS	70
4.3. FUTURE RESEARCH	71
4.4. CONCLUSIONS.....	72
BIBLIOGRAPHY	73
ANNEX A. Exploratory analysis with NC collected in 2014 for the ET study.....	85
ANNEX B. Assessment of overdispersion for NC	95
ANNEX C. Results with NC collected in December 2016 and comparison with 2014 results	96
ANNEX D. Estimation of the intervention effect on non-citable articles using zero-inflated models	104
ANNEX E. Comparison of sample density and theoretical densities of Poisson, QP, ZIP, NB and ZINB	105
ANNEX F. Characteristics of recurrent events data models.....	106
ANNEX G. Simulation study to compare shared frailty models for recurrent events.....	107
ANNEX H. Results of the adjusted Frailty models for recurrent events	111
ANNEX I. Main results with NC collected in July 2020.....	112
ANNEX J. Article published in BMC Medical Research Methodology	115
ANNEX K. R code for the analyses	122

LIST OF FIGURES

Figure 1. Overview of the editorial peer review process for indexed journals	14
Figure 2. Summary of peer reviewer involvement when evaluating a manuscript	16
Figure 3. Percentage of unique and overlapping citations in Google Scholar, Web of Science and Scopus (n=2,448,055 citations of the 2,515 highly-cited documents written in English and published in 2006; from the 2017 edition of the Google Scholar Classic Papers list.)	22
Figure 4. Scheme of the allocation of interventions of the IQ study. R=reference; C=Checklist; S=Statistician; SC= both Statistician and Checklist.	25
Figure 5. Summary of the flow chart of the IQ study. R=reference; C=Checklist; S=Statistician; SC= both Statistician and Checklist.....	26
Figure 6. Scheme of the allocation of interventions for the ET study. R=Reference; SC= both Statistician and Checklist.	27
Figure 7. Summary of the flow chart of the ET study. R=Reference; SC= both Statistician and Checklist.	27
Figure 8. Scheme of the data collection evolution. NC=Number of citations; R=Reference group; SC=Intervention group with the additional review of a statistician using reporting guidelines; SD=Standard deviation	32
Figure 9. Comparison of the count models for the type of data to interpret	41
Figure 10. Steps towards reproducibility	51
Figure 11. Number of citations by study and intervention group	54
Figure 12. Evolution of citations by group in the ET study	56
Figure 13. Mean cumulative function for the ET study articles.....	56
Figure 14. Citations-per-year mean ratio for the main comparison (R vs. SC) using the Poisson model with Wald standard errors	57
Figure 15. Citations-per-year mean ratio for the main comparison (R vs. SC)	59
Figure 16. Citation-per-year mean ratio using QP, NB, ZIP and ZINB models.....	60
Figure 17. Empirical and theoretical CDFs for the ET study. For clarity, this figure shows two plots: 17a for the CDFs of Poisson, QP and ZIP, and 17b for NB and ZINB	61

Figure 18. Citations-per-year mean ratio of SC additional intervention by study design63

Figure 19. Mean cumulative function for the ET study by group64

Figure 20. Mean cumulative function differences between groups for the ET study65

Figure 21. Comparison of the expected number of citations at a certain time between reference and intervention groups for ET and IQ studies66

Figure 22. Gamma density of the estimated frailty distribution67

Figure 23. Citations-per-year mean ratio for Statistician and Reporting Guidelines effects68

LIST OF TABLES

Table 1. Reporting guidelines (RG) for main study designs	18
Table 2. Syntax of the search details for the different data sources	23
Table 3. Description of the common variables for the Articles and Citations datasets, as well as for the extra variables belonging to the Citations dataset	31
Table 4. R packages and functions for fitting shared frailty models for recurrent events.....	50
Table 5. Names of the variables and their meanings that appear in the first tab of the shiny app.	52
Table 6. Number of citations by study and intervention group until December 2018. Groups not included in the main analysis are in a shaded style.....	55
Table 7. Kolmogorov-Smirnov statistic by study and model.....	61
Table 8. Summary of the adjusted by baseline MQAI intervention effect by different count models for ET study	62
Table 9. Summary of the adjusted by baseline MQAI intervention effect by different count models for IQ study	62
Table 9. Estimation of the hazard ratio comparing the expected number of citations between R and SC groups for ET and IQ studies.	66

ABBREVIATIONS

AE: Associate editor

AIC: Akaike Information Criterion

ASA: American Statistical Association

BIC: Bayesian Information Criterion

CENTRAL: Cochrane Central Register of Controlled Trials

C: Intervention group with suggestion to the reviewers to employ a local Checklist

CI: Confidence interval

EIC: Editor-in-Chief

ET: “Enhance Transparency” study. Cobo E, Selva-O’Callaghan A, Ribera JM, Cardellach F, Dominguez R, Vilardell M. Statistical reviewers improve reporting in Biomedical articles: a randomized trial. *Plos One*. 2007; 2 (3): e332.

GLM: Generalized Linear Model

GS: Google Scholar

HR: Hazard ratio

IF: Index factor

IG: Inverse Gaussian

IQ: “Improve Quality” study. Cobo E, Cortés J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomized trial. *BMJ*. 2011; 343: d6783.

JK: Jackknife method

K-S: Kolmogorov-Smirnov

LN: Log-normal

MA: Meta-analysis methodology

MCF: Mean cumulative function

MQAI: Manuscript Quality Assessment Instrument

ML: Maximum likelihood

MML: Maximum marginal likelihood

NB: Negative Binomial

NC: Number of citations

NLS: Nonlinear least squares

PPL: Penalized partial likelihood

PS: Positive stable

PVF: Power variance function

Q1: First quartile

Q3: Third quartile

qAIC: quasi Akaike Information Criterion

QP: Quasi-Poisson

R: Reference group, traditional review

RCTs: Randomized controlled trials

RG: Reporting guidelines

S: Intervention group adding a Statistician from the reviewer panel list

SC: Intervention group adding a Statistician from the reviewer panel list using reporting guidelines

SD: Standard Deviation

WHO: World health organization

WoS: Web of Science

ZINB: Zero-Inflated Negative Binomial

ZIP: Zero-Inflated Poisson model

Chapter 1

INTRODUCTION

1.1. GENERAL INFORMATION

This doctoral thesis relates to the importance of proposing interventions capable of improving the quality of biomedical research as well as its impact and dissemination. The growing need to conduct research on research motivates this work. Evaluating whether the research produced is of quality and has a positive impact on the development of knowledge is fundamental, and this is even truer in the field of biomedical research, where the health of people is at stake.

Biomedical research has an unrestricted potential to improve the quantity and quality of life. It is a necessary activity for the success of any strategy that aims to improve health (1). The integration of research with clinical practice promotes higher quality of health services and better and faster implementation of scientific advances in prevention, diagnosis, patterns of care, treatment of diseases, costs and use. In addition, health policies and public health services will be better informed and oriented when responding to the needs and demands of citizens (2) (3).

In 2009, Chalmers and Glasziou (4) estimated that 85% of all research funding was wasted due to inappropriate research questions, flawed execution, inappropriate study design, irrelevant endpoints, poor reporting and/or non-publication (5). In 2014, The Lancet published a series of five papers about research discussing the abovementioned weaknesses (6) (7) (8) (9) (10). Moher et al. (11) published a review in 2016 showing gratifying action in response to the weaknesses discussed on The Lancet's series, although they claimed that much more needs to be done to reduce research

waste. One of the strategies proposed for controlling this waste involves getting journal editors to promote the use of reporting guidelines (RG) by peer reviewers and authors.

A critical step in scientific production is the review and editorial processes during the publishing period, as it is essentially a quality control mechanism (12). Editorial peer review has been widely accepted as the best way to filter low quality research and to improve papers with significant contributions to their fields (13) (14), although some “studies showed that peer reviewers were not able to appropriately detect errors, improve the completeness of reporting, or decrease the distortion of the study results” (15). A systematic review performed in 2005 and published in 2007 by Jefferson et al. (16) evaluated the effect of editorial peer review processes in any type of design, although they focused mainly on observational studies. They concluded that there was little empirical evidence available to support the use of editorial peer review as a mechanism to ensure quality in biomedical publications.

Transparent and complete reporting is essential for allowing readers to assess the validity of research findings for use in healthcare and policy decision-making (17). To this end, and in an attempt to address problems in reporting health research, researchers and methodologists have collaborated with journal editors to develop guidelines as a quality assurance tool for complete and transparent reporting of health research results (18). After the development of the first reporting guidelines for randomized controlled trials (RCTs) in 1996 (19), other guidelines have come into regular use and new ones are continually being developed in a systematic way (20).

Despite all the efforts, many RGs are underused (17). In order to increase their uptake, evidence of their effectiveness is required. RGs would provide authors, peer reviewers and editors with an important resource for the use and implementation of pertinent guidance (17).

When we talk about the progress of science, what we look for is to quantify the impact that current research has on subsequent science. For that, we have to guarantee not only the quality of the manuscripts, but also its long-term impact. It is expected that the quality of the manuscripts is related to this impact. Citation analysis is a traditional method for assessing research impact, and it is performed by examining an individual publication and assessing how often it has been cited, if ever, by subsequent publications. It is a tool for gauging the extent of a publication’s influence in the literature and for tracking the advancement of knowledge with the inherent assumption that significant publications will demonstrate a high citation count. Citation analysis has some well-known and accepted flaws, such as self-citing, reciprocal citing by colleagues and inclusion of “negative” citations. However, it is accepted as a standard tool for assessing the merits of a publication (21).

After a bibliographic search, we found two studies that randomly allocated the interventions of including a statistical reviewer and/or publication guidelines during the peer review process with the aim of improving manuscript quality. To observe the long-term effect on the impact of these interventions, we have had the chance to collect posterior new information: the number of citations (NC) for all the articles included in the two randomized studies.

In studying the effects on impact as a result of including a statistical reviewer and/or publication guidelines during the peer review process, the traditional way would have been to size the study with some pre-desired power to test this hypothesis. In order to design such a confirmatory study, we would have needed information about the distribution of this variable among the two groups under both scenarios: the null and alternative hypotheses. In classical pharmacological situations, the outcome would be represented by a Normal homoscedastic distribution with the same mean under the null hypothesis, but the alternative hypothesis would have a separation equal to the expected effect size, which is referred to as the delta. However, as there are no previous studies for such intervention, we do not have this information. Therefore, we have done a proof of concept study that aims to assess these long-term interventions in order to have an idea of the expected effect as well as their viability at a confirmation level.

Apart from the total NC an article receives, we also have the time between citations; thus, we also analyzed the NC with recurrent events methods. We think that taking advantage of all the collected information can give a more realistic estimate of the effect.

Recurrent events are repeated events of the same type, such as acute exacerbations in asthmatic children, seizures in epileptics, myocardial infarctions, migraine attacks and sports injuries.

In a systematic review, Meghan et al. (22) revealed that an appropriate statistical method was used in less than one-third of 83 research articles whose outcome of interest was a recurrent event. Any application of sub-optimal suitability could lead to loss in terms of bias and the precision of the results.

Whenever information on time is collected throughout the study, and when information on event time plays an important role in addressing the true research question, survival techniques are always a better choice than non-survival techniques. For example, one may be interested in knowing whether the intervention is responsible for increasing time between successive events (22).

In this context, those survival times observed in the same individual will not be independent of each other due to the existence of individual heterogeneity or within-subject dependence. Furthermore, the effects of the explanatory variables may be different between the first event (or recurrence) and the next.

1.2. DISSERTATION STRUCTURE

The organization of this thesis follows a commonly used structure in biomedical research. There are four chapters: Introduction, Methods, Results and Discussion.

In the Introduction, you will find the state of existing research on several concepts related to the thesis objectives. The aim is to establish a background for understanding all the processes. This chapter details how the editorial peer review process has evolved over the years and its importance within research. It also explains: the use and evaluation of RGs as a tool for improving manuscripts; and the indicators used to measure and evaluate research impact. There, we explain the extension of our search for a previous systematic review (15) in order to summarize all published interventions thought to improve manuscripts throughout editorial peer review in biomedical research journals. We introduce two previous studies (ET and IQ studies) conducted from 2004 to 2005 and 2009 to 2010, which randomly allocated two interventions: adding a statistician and using RG in the editorial peer review panel of the journal *Medicina Clínica* (Barcelona). Both studies aimed to see the effect of these interventions on MQAI.

Chapter 2 concerns the methodology used to analyze the data. It introduces the data collection process of the NC for the articles previously randomized in the ET and IQ studies, as well as other variables used for the analyses. Before seeing the distribution of NC, we planned to analyze this variable with the most popular model for count data, the Poisson model. After collecting the data and seeing its distribution, we realized that the Poisson model, which was defined a priori, was not the most appropriate model for fitting the NC. This is because, despite the Poisson model equating mean and variance, the NC for the latter was much larger than the former. This implies overdispersion, and the Poisson model does not estimate standard errors correctly. The estimation we use is free of model assumptions and is based on the Jackknife method (JK) for the main analysis. In order to see if the findings are consistent among different statistical methods, a sensitivity analysis uses other count models more appropriate for fitting count data with overdispersion, such as the Quasi-Poisson (QP), the Negative Binomial (NB), the Zero-Inflated Poisson (ZIP) and the Zero-Inflated Negative Binomial (ZINB).

Once count data models are presented, we analyze NC with recurrent events methods. The second part of the methodology in Chapter 2 defines the models used in the multivariate time-to-event analysis. First, we introduce the non-parametric mean cumulative function (MCF) method to describe the average NC that occurs for one individual within a certain time, and we compare the intensity of citations between groups. Then, we examine frailty models, which are a kind of random effects model for survival analysis. These models estimate the dependence between citations while explicitly specifying its probability distribution and including it in the model as a frailty parameter.

Chapter 3 presents the results and describes both the NC and our analysis of the effect of the interventions. For the count data models and recurrent events methods, we present the effects of the three interventions: (1) statistical reviewer using RG; (2) the statistical reviewer alone; and (3) the RG alone.

Finally, Chapter 4 discusses the results obtained, the limitations of the study and proposes interesting new lines of research.

1.3. STATE OF EXISTING RESEARCH

1.3.1. EDITORIAL PEER REVIEW PROCESS

Biomedical journals are a tool for the dissemination and critical evaluation of advances in biomedical research. Nowadays, the editorial process for journals is based on peer review. The International Committee of Medical Journal Editors defines editorial peer review as the “critical assessment of manuscripts submitted to journals by experts who are not part of the editorial staff” (23).

The editorial peer review process seeks to ensure the quality of both content and methodology in submitted manuscripts, with two basic objectives (16) (24):

- Select the best manuscripts in the literature and filter out the manuscripts that do not meet the minimum quality requirements.
- Improve through recommendations and guidelines the methodological quality and the communication of results.

Thus, peer-reviewed journals are the most important tool for selecting and improving papers to be published in biomedical research (25). As such, we could consider that the progress of science relies on the editorial peer review process.

Although it can vary between journals, the editorial peer review process can be broadly summarized in 12 steps (26):

1. Submission of paper: Author submits the paper to the journal.
2. Editorial office assessment: The journal checks the paper’s composition and arrangements against the journal’s Author Guidelines to make sure it includes the required sections and stylizations.
3. Appraisal by the Editor-in-Chief (EIC): The EIC checks that the paper is appropriate for the journal and is sufficiently original and interesting. If not, the paper may be rejected without being reviewed any further.
4. EIC assigns an associate editor (AE): Some journals have associate editors who handle the peer review. If they do, they would be assigned at this stage to act as “handling editor”.
5. Invitation to reviewers: The handling editor sends invitations to individuals he or she believes would be appropriate reviewers.
6. Response to invitations: Potential reviewers consider the invitation against their own expertise, conflicts of interest and availability. Then they accept or decline.

7. Review is conducted: The reviewer reads the paper, ideally several times, and takes notes so as to build a detailed point-by-point review. The review is then submitted to the journal with a recommendation to accept or reject it – or else with a request for revision (usually flagged as either major or minor) before it is reconsidered.
8. The journal evaluates the reviews: The handling editor considers all the returned reviews before making an overall decision.
9. The decision is communicated: The handling editor sends a decision (accept/reject/revision) email to the author, including any relevant reviewer comments.
10. If the article is rejected or sent back for either major or minor revision, the handling editor includes constructive comments from the reviewers to help the author improve the article. If the paper was sent back for revision, the reviewers may receive a new version.
11. If authors accept the review, they re-start the process with a new re-submission.
12. Next steps: If accepted, the paper is sent to production.

Figure 1 shows an overview of the editorial peer review process (26), where you can identify the different steps. Some of the journals with a higher index factor (IF) repeat these steps more than once.

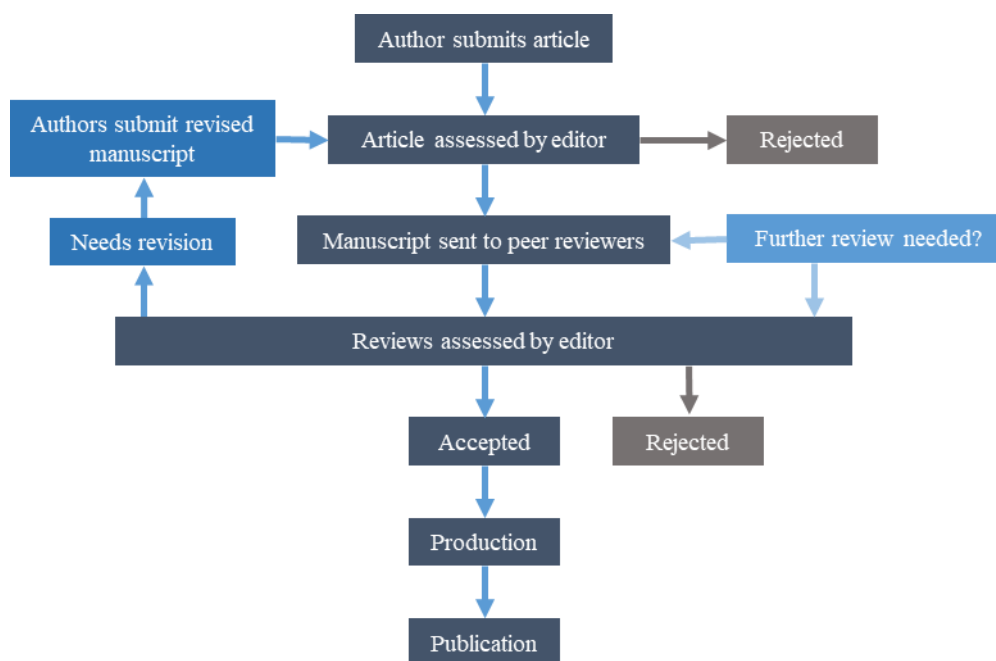


Figure 1. Overview of the editorial peer review process for indexed journals

Given the importance of the editorial peer review process, different tools have been used in biomedical research for assessing the quality of peer review reports (27). A systematic review

published in 2009 (27) identified 24 tools: 23 scales and 1 checklist. One out of four tools consisted of a single item that simply asked the evaluator for a direct assessment of the peer review report's "overall quality". The remaining tools had between 4 to 26 items in which the overall quality was assessed as the sum of all items, their mean, or as a summary score.

Sense about Science (28) surveyed 3000 researchers to ascertain their current view of editorial peer review. They noticed that 90% of respondents felt that editorial peer review improves the quality of research articles overall, and 85% thought that without editorial peer review there is no control in scientific communication. When respondents were asked for the most important thing for maintaining a healthy editorial peer review system, 66% felt that reviewers should be given clear guidance on the criteria for reviewing, and 37% opted for formal training of reviewers to create a quality benchmark.

They commented that editors and peer reviewers in biomedical journals lack the appropriate competencies. Peer reviewers rarely receive formal training (24) (29), and thus their capacity to detect errors and identify deficiencies in reporting is limited (30). Facing this deficit in editorial peer review training, some researchers described general and specific elements that should be included in a high-quality review for the journal (26) (31). Figure 2 shows a summary of peer reviewer involvement in the revision process, sorted by importance according to Kelly et al. (26).

The benefit of the editorial peer review process has been questioned (24) (32) (33) (34) (35) (36). Not only are there gaps in the part of the process that includes reviewers, but editorial staffs also demonstrate poor practices in many cases (37) (38) (39). In addition, despite its wide acceptance by the scientific community, little is known about its impact on the quality of reporting the published research (40).

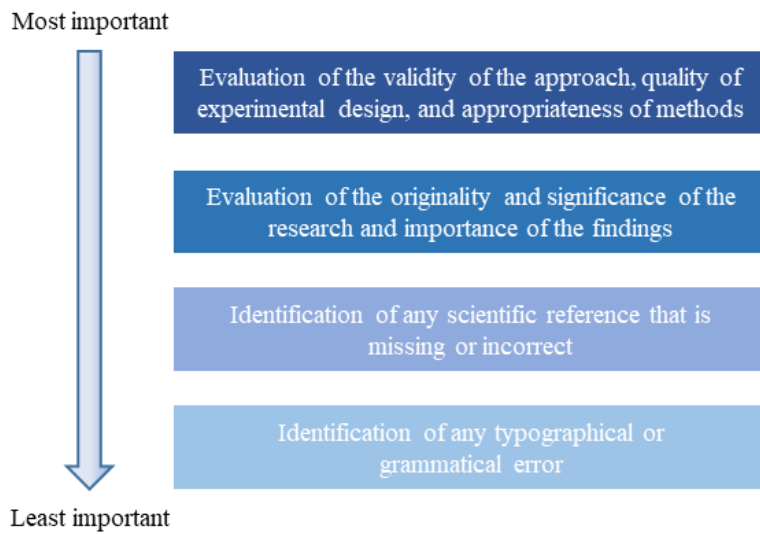


Figure 2. Summary of peer reviewer involvement when evaluating a manuscript

One of the limitations of the editorial peer review process is the lack of training and specialization of peer reviewers, which appears to compromise the scientific quality of published manuscripts (24) (40). Current editors of scientific journals are expected to monitor the thoroughness and fairness of the editorial peer review process (41). Although statistical aspects play an important role in medical research, in many cases the published research lacks the expected rigor (42) (43). To avoid statistical pitfalls in the review process (44) (45), some journals make available to the authors guidelines to communicate the description of data, expectations for statistical usage, and the preparation of findings (46) (47); but most do not (37). Enhanced expert statistical reviews of manuscripts would help confront a key part of the peer review evaluation. It would allow recognizing the weaknesses in the design and statistical analyses in order to avoid flaws that either: 1) limit or preclude the potential contributions of findings to the literature or 2) create misleading results (48) (49).

1.3.2. REPORTING GUIDELINES

RGs have been developed over the past two decades to provide advice on how to report research methods and findings; this helps authors, editors and peer reviewers check and improve the transparency of research studies while ensuring that papers are both accurate and complete (50) (51) (52) (53) (54) (55). RGs are usually in the form of a checklist, a flow diagram or an explanation and elaboration document. They specify a minimum set of items required for a clear, transparent account of what was done and what was found in a research study. A large number of RGs have been published during the last 15 years. They are heterogeneous in the sense that each of them has different characteristics in terms of their scope, development methodology and presentation of recommendations.

Since the development in 1996 of the CONSolidated Standards Of Reporting Trials (CONSORT) statement for reporting randomized controlled trials, several guidelines have been developed relating to other types of research (56). Table 1 presents the RGs for the main study designs (57).

An international initiative promoting transparent and accurate reporting of health-related research was launched in 2008: the EQUATOR (Enhancing the QUALity and Transparency of health Research) Network. This program supports wider practical implementation of RGs by all relevant parties to increase the usability and value of health research.

Many health journals endorse RG (in particular, the CONSORT Statement, endorsed by over 600 journals) (55); thus, their contributing authors should follow them. Those journals usually include a statement in the “Instructions to Authors” suggesting or requiring the use of one or more guidelines to submit a manuscript. However, there is still skepticism. A recent study (58) identified 257 journals that publish observational studies (2007-2017) related to the scope of seven Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) extensions (antimicrobial stewardship programs, infectious disease molecular epidemiology, molecular epidemiology, rheumatology, genetic association studies, routinely collected health data, and response-driven sampling). Of these 257 journals, 12 (5%) required the STROBE statement on submission; 22 (9%) suggested using them; 12 (5%) recommended a “relevant guideline”; 72 (28%) mentioned it indirectly (via editorial policies or International Committee of Medical Journal Editors recommendations); and 139 (54%) did not mention STROBE. They concluded that RG endorsement rates are low. Another study (59) surveyed the authors of observational studies to know their attitude toward and experiences with STROBE. Of the 1015 non-randomly sampled participants who completed the survey, 185 (18.2%) indicated they had never heard of STROBE nor used it previously; 195 (19.2%) had heard of it but never used it; and 635 (62.6%) had used it.

Type of study	RG Acronym	RG large name
Randomized trials	CONSORT	Consolidated Standards of Reporting Trials (60)
Observational studies	STROBE	Strengthening the Reporting of Observational Studies in Epidemiology (61)
Systematic reviews	PRISMA	Transparent reporting of systematic reviews and meta-analysis (62)
Study protocols	SPIRIT	Defining standard protocol items for clinical trials (63)
	PRISMA-P	Preferred Reporting Items for Systematic review and Meta-Analysis Protocols (64)
Diagnostic/Prognostic studies	STARD	Standards for Reporting of Diagnostic Accuracy Studies (65)
	TRIPOD	Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (66)
Case reports	CARE	Consensus-based clinical case reporting guideline development (67)
Clinical practice guidelines	AGREE	The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines (68)
	RIGHT	A Reporting Tool for Practice Guidelines in Health Care (69)
Qualitative research	SRQR	Standards for reporting qualitative research: a synthesis of recommendations (70)
	COREQ	Consolidated criteria for reporting qualitative research (71)
Animal pre-clinical studies	ARRIVE	Animal Research: Reporting In Vivo Experiments Guidelines (72)
Quality improvement studies	SQUIRE	SQUIRE 2.0 (Standards for QUality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process (73)
Economic evaluations	CHEERS	Consolidated health economic evaluation reporting standards statement (74)

Table 1. Reporting guidelines (RG) for main study designs.

To date, there have been few evaluations of RGs. One reporting guideline extensively evaluated is the CONSORT Statement (18) (19) (75) (76) (77). In 2006, a systematic review of CONSORT effectiveness identified eight studies evaluating its impact (78). A 2012 update of that review identified 42 additional evaluations (60) (79). Both found that endorsement of CONSORT by journals is significantly associated with a higher frequency of completely reported trials, at least for some items of the CONSORT Statement.

There is growing evidence that use of a checklist is a beneficial tool. For example, Richards (80) concluded in his systematic review,

“There is evidence that [checklists for reporting a range of study designs have] had an impact on the quality of reporting clinical trials, at least. These checklists can assist both authors and peer reviewers to improve the quality of publications and, together with the additional research into the peer review process called for by the authors here, should lead to improvements in the quality of scientific publications in the future.”

However, although initial studies evaluating the impact of journal support for RG indicate beneficial effects on the completeness and transparency of publications (51), better use and adherence of reporting checklists by journal editors, peer reviewers, and authors could be one important step towards improving the reporting of published articles. Improving adherence to RGs is one of the key issues for enhancing complete and accurate reporting and therefore reduce waste in research (81).

Blanco et al. (81) reveal that, “most published research aimed at improving adherence to RGs has been conducted in journals”. They summarized journal strategies that range from “making available editorial statements that endorse certain RGs, recommending or requiring authors to follow RGs in the ‘Instructions to authors’, and requiring authors to submit an RG checklist together with the manuscript, with page numbers indicated for each item”. However, even though some journals have implemented these strategies, there is no sufficient research on their effect.

1.3.3. RESEARCH IMPACT

“Research impact is the demonstrable contribution that research makes to the economy, society, environment and culture beyond the contribution to academic research”¹

We consider the impact of a scientific article on research to be the attention it gets, and this attention can be assessed using citation analysis. Citation by other authors is important in the dissemination of published study findings. Citations represent the author’s explicit recognition of having relied on previous knowledge to build new knowledge (82).

In this context, Egghe & Rousseau (83) claim four important assumptions:

- (1) A citation of an article implies use of that document by the citing author.
- (2) A citation reflects the merit (quality, significance, impact) of the article.
- (3) References are made to the best possible works.
- (4) An article is related in content to the one in which it is cited.

Thus, NC can be regarded as an objective method of obtaining a quantitative expression of the utilization and contribution of a particular published paper, which could be interpreted as the impact of a paper (38). However, a critical issue is whether or not citation is adequate for describing the quality and impact of research. One of the main factors related to NC is the quality of the paper, which could predict the paper’s future citation impact (84) (85) (86) (87) (88) (89). Some studies have mentioned that papers with higher quality obtain more citations (84) (87). According to Garfield (90), the Pareto principle applies, as about 20% of papers receive 80% of the total NC, while 80% of papers are rarely cited or not cited at all. It is commonly accepted that the top 20% of papers have higher quality compared to other papers (89). Therefore, the quality of a paper is a factor that influences the impact of research.

Peer-reviewed papers obtain more citations than non-reviewed ones (91). Aside from the quality of a paper, some articles indicate that among the factors influencing citations are the novelties of its subject, the popularity of the topics, and the citer’s interest in the subject. The initial citations that a paper receives constitute the early feedback of the scientific community about that paper. The NC a paper receives in the years immediately following its publication can be considered as a predictor of its future citations (92) (93). On the other hand, as stated by Garner et al. “papers that get cited further away will tend to take longer to pick up cites” (94).

However, citation counts are not free of limitations, since there are many reasons why researchers cite others’ works in their papers. For instance, they cite others to support their own claims, methodology or

¹ Australian Research Council. <https://www.arc.gov.au/>. [Online] 2017.

findings (called supportive citations). Other citations are made to present other researchers' points of view. Some papers are cited as good research examples, while others might be cited as a negative example of poorly designed research (95) (96).

A different measure that applies to the journal instead of the paper is the impact factor (IF). IF is defined as the ratio between citations and recent citable items published. Thus, the IF of a journal is calculated by taking the number of current year citations to the source items published in that journal during the previous two years and dividing that by the total number of such source items (97). For example, the 2018 IF for the *Cancer Journal for Clinicians* was 72.58, the highest among all scientific journals (98). This number is calculated by noting that 43 source items were published in 2016 and 43 items in 2017; in turn, the journal's 2016 and 2017 material was cited a total of 6241 times in 2018 ($6241/86 = 72.58$). Journals with higher IF are deemed to be more important than those with lower ones. As papers are nested within the publishing journal, IF is a related measure. This measure has a strong relationship with the NC. In an exploratory way, I considered the sum of IF of the journals where the citing paper was published as a secondary measure of impact in the 2014 analyses.

All the articles that set up the ET and IQ studies were submitted to the same journal (*Medicina Clínica*, Barcelona); thus, they all have the same assessment of the IF of the journal where they were published.

A criticism of NC is the time it takes to appear. With current communication speeds, researchers, authors, and grant funders are impatient to get an indicator of its value. Waiting 1-3 years for publication and citation seems interminable.

In recent years, there has been an explosion in the use of social networks. These networks have extended the dissemination of information and have led to the appearance of new indicators of research activity, known as "alternative metrics" or altmetrics (99).

Sense about Science (28) published a survey finding that when a reader is interacting with research outputs (e.g., research articles, preprints, data), the additional information that respondents found helpful for assessing those research outputs are citations (88%). Respondents view downloads helpful (69%) and notice the number of retweets and/or mentions on blogs (31%).

Some studies (100) (101) have evaluated the association between altmetric scores and citation indicators. The findings indicate a positive correlation (Pearson correlation coefficient = 0.462) between the number of citations and the altmetric score of articles. Replacing NC with this social score is not recommended, but it is possible to use altmetric indicators not as an alternative to traditional citation bibliometrics but as useful complementary information (99).

Altmetrics started tracking attention to research across various sources (Twitter, Facebook, news, blogs, Mendeley, etc.) in October 2011 (82); therefore, we cannot expand our work using altmetric scores, since the articles in the ET and IQ studies were published before that date.

There are several citation databases of peer-reviewed literature for collecting NC, like Google Scholar (GS), Web of Science (WoS) and Scopus. All deliver a comprehensive overview of the world's research output in the fields of science, technology, medicine, social sciences, etc. A recent study (102) presented the results of a comprehensive analysis of these citation sources (Figure 3). They identify that 46.9% of citations were found by all three sources. GS found over a third (36.9%) of all citations. WoS and/or Scopus found only 6% of all citations that were not found by GS.

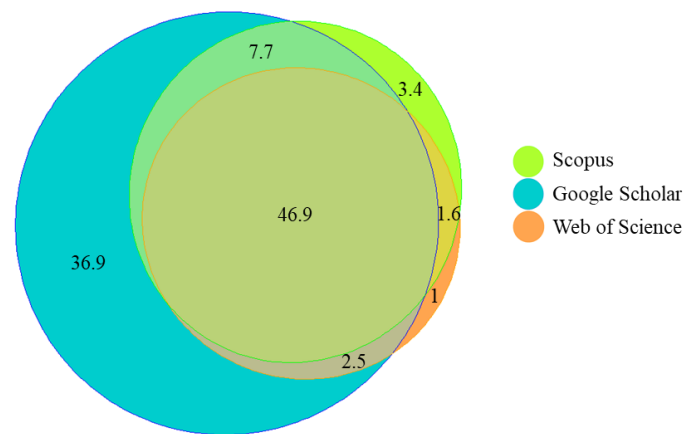


Figure 3. Percentage of unique and overlapping citations in Google Scholar, Web of Science and Scopus (n=2,448,055 citations of the 2,515 highly-cited documents written in English and published in 2006; from the 2017 edition of the Google Scholar Classic Papers list.²)

The Spearman correlation between citation counts of GS and WoS was 0.98 for the Biomedical Sciences category, and the mean ratio of citation counts (GS / WoS) was 1.72, indicating that the mean number of citations in GS is 72% higher than for WoS. The authors do not provide confidence intervals (CI) of the estimations.

² https://scholar.google.com/citations?view_op=list_classic_articles&hl=en&by=2006

1.3.5. INTERVENTIONS TO IMPROVE MANUSCRIPTS THROUGHOUT EDITORIAL PEER REVIEW IN BIOMEDICAL JOURNALS

In searching the literature for interventions implemented to improve editorial peer review in biomedical research, we found a review published in 2015 by Bruce et al. (15) with this very objective. Therefore, a good strategy to see if there are new studies is to expand this review with a longer follow-up: the period between 15 Jun 2015 and 1 May 2017.

Just Bruce et al. (15) did, we searched the Cochrane Central Register of Controlled Trials (CENTRAL), MEDLINE (via PubMed), Embase, and the WHO International Clinical Trials Registry Platform for all reports of RCTs evaluating the impact of interventions aiming to improve the quality of editorial peer review in biomedical journals. We also searched the Cochrane Database of Systematic Reviews to identify systematic reviews on the editorial peer review process. We placed no limitations on language. The search strategy relied on the Cochrane Highly Sensitive Search Strategies (103) and the search term “peer review”. Table 2 shows the syntax of the search details for the different data sources. We also hand-searched reference lists of reports and reviews that were identified during the screening process and known to be dedicated to the editorial peer review process.

Data source	Syntax
Pubmed and Embase	("peer review"[All Fields] AND ("2015/06/15"[PDAT] : "2017/05/01"[PDAT])) AND Randomized Controlled Trial[Publication Type]
WHO International clinical trials registry platform and Cochrane database of systematic reviews	*peer review [Complete text] AND randomized controlled trial [Type of publication] – with date of publication between Jun 2015 and May 2017

Table 2. Syntax of the search details for the different data sources

Bruce et al. (15) pre-specified the categorization of the interventions as follows:

- Training, which included training or mentoring programs for peer reviewers to provide instructional support for appropriately evaluating important components of manuscript submissions
- Addition of peer reviewers for specific tasks, such as adding a statistical peer reviewer
- Peer reviewers' use of a checklist, such as reporting guideline checklists
- “Open” peer review process
- “Blinded”/masked peer review
- Other interventions

They selected 22 reports of randomized controlled trials, in which they found 25 comparisons evaluating training interventions (n=5), the addition of a statistical peer reviewer (n=2), use of a checklist (n=2), open peer review (n=7), blinded peer review (n=6) and other interventions to increase the speed of the editorial peer review process (n=3).

They defined different outcomes to measure the impact of the interventions described above:

- Final manuscript quality. They found several scales to assess manuscript quality: MQAI and other scales (76) that measure the completeness of reporting based on reporting guideline items.
- Quality of the peer review report, measured by scales such as the Review Quality Instrument (104) or editor routine quality rating scales (105).
- Rejection rate.
- Time spent on the peer review as reported by peer reviewers.
- Overall duration of the peer review process.

The most-used outcome was the quality of the peer review report (n=14), followed by the rejection rate (n=10), the time spent on the peer review as reported by peer review (n=7), the final manuscript quality (n=4) and the overall duration of the peer review process (n=2). Ten of the 25 comparisons included more than one outcome. The four comparisons using the final manuscript quality outcome were made in three publications.

With the expansion of the review, we found one additional trial (106) that studied the effect of the “blinded”/masked peer review intervention on the outcome rejection rate.

Therefore, at least 23 randomized trials have studied interventions during the editorial peer review process. The majority of them focus on the quality of editorial peer review as a surrogate outcome while only three (107) (108) (109) analyzed the completeness of reporting as an outcome. Of these three trials, two (108) (107) were based on a single journal (Medicina Clinica; Elsevier, Barcelona) and they found partially positive results regarding paper completeness. Both interventions were based on the use of methodological reviewers and RG. Those studies were based on a partly subjective outcome, MQAI, and there was some evidence that evaluators could have successfully guessed which ones were in the intervention group (15), since masked evaluators were able to guess the allocated group in 62% (56/90) of the papers, thus presenting a risk of detection bias. We follow up on those studies here by taking advantage of the WoS (110) to reassess those two trials by using the NC later received by those papers.

One of the two trials identified was published in 2007 (107). The editorial team of Medicina Clinica published the results of a masked, randomized experiment to improve the effect of editorial peer review on paper quality as assessed by the MQAI. We refer to this work as the IQ study. The two crossed interventions were (1) adding a peer reviewer from the standard list of methodological reviewers and (2) suggesting an RG checklist to reviewers. They randomly allocated the provisionally accepted

manuscripts into four groups defined by the orthogonally tested interventions: clinical reviewers (R) as the normal reference procedure; clinical reviewers plus a statistical reviewer (S); clinical reviewers with checklist (C); and clinical reviewers plus a statistical reviewer and checklist (SC). Figure 4 shows the scheme of the allocation of interventions for the IQ study. The checklist intervention (C) consisted of just one recommendation:

“To facilitate your revision, you will find enclosed the reporting guideline from Bosch and Guardiola (Med Clin (Barc) 2003; 121:228–30) (111). If you prefer, you may also employ one of the following documents: for clinical trials, the CONSORT statement (Ann Inter Med 2001;134:663–694) (60); for meta-analysis, QUOROM (Lancet 1999; 354: 1896–1900) (112); for diagnostic tests, STARD (Clinical Chemistry 2003;49:7–18) (65); or the collections provided by the Scottish Intercollegiate Guidelines Network (<http://www.sign.ac.uk/guidelines/fulltext/50/annexc.html>) or by Mora (Med Clin (Barc) 1999;113: 138–49) (113)”

The statistical reviewer (S) intervention comprised an additional review of the manuscript by a statistical reviewer chosen from Medicina Clínica’s traditional list of experts. They found positive evidence for the effects of adding a methodological expert to the review panel.

IQ study		Checklist	
		No	Yes
Statistician	No	R	C
	Yes	S	SC

Figure 4. Scheme of the allocation of interventions of the IQ study. *R*=reference; *C*=Checklist; *S*=Statistician; *SC*= both Statistician and Checklist.

Figure 5 shows the flow chart of the IQ study. Of the 327 original articles received between May 2004 and March 2005, 196 (59.9%) were directly rejected by the editorial team. The remaining 131 (40.1%) were selected by the editorial committee as possible publications and therefore randomized. Of these, two were excluded either as a result of an administrative error ($n = 1$) or because the authors refused to participate ($n = 1$).

Of the 129 randomized manuscripts:

- 14 were lost to follow-up because authors missed the deadline and the masked allocation was revealed.
- 21 (18.3%) of the 115 included papers related to “interventional” studies, but only 3 were randomized clinical trials.
- 46 (40.0%) were longitudinal designs.
- 28 (24.3%) were cross-sectional.
- 20 (17.4%) others.
- 16 manuscripts were rejected by the editorial team after evaluating peer review reports

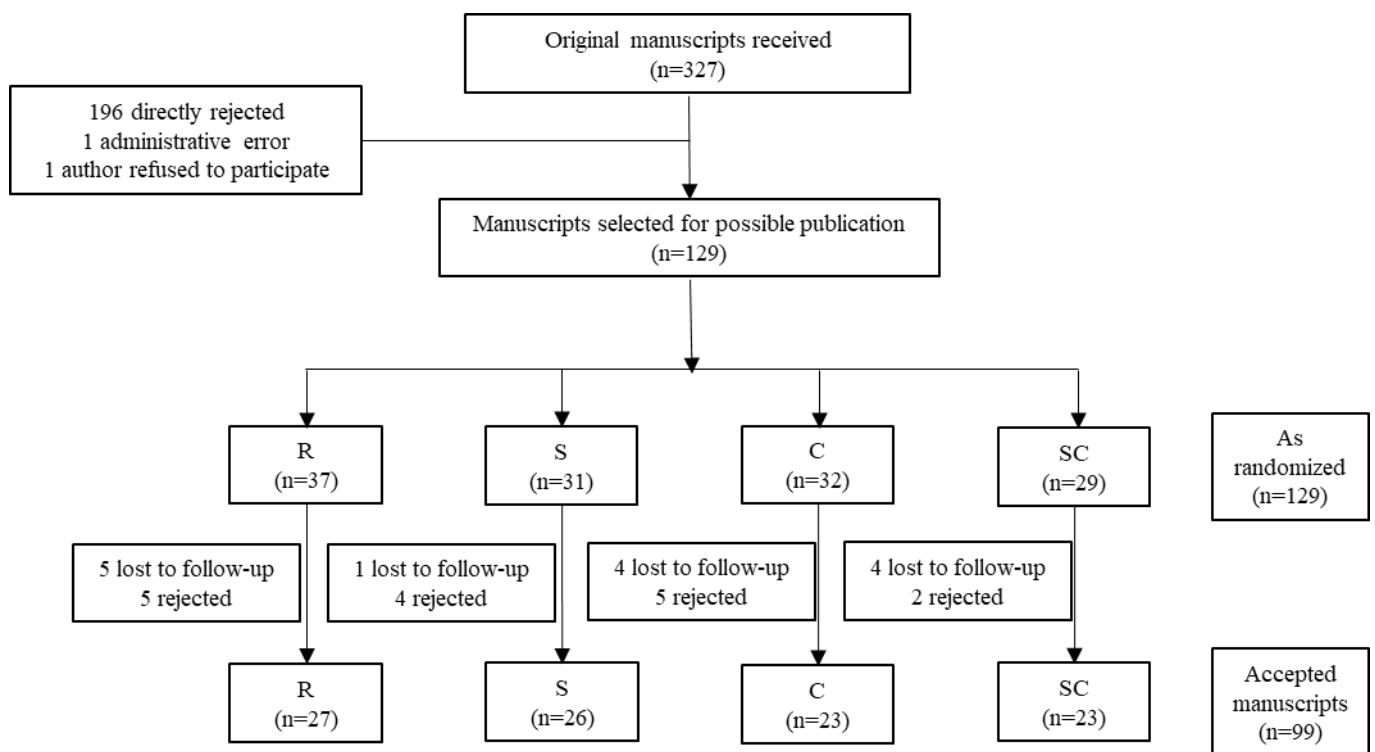


Figure 5. Summary of the flow chart of the IQ study. *R=reference; C=Checklist; S=Statistician; SC= both Statistician and Checklist.*

Due to the small effect observed in the IQ study, the same authors published a new study that redefined and pooled both interventions into a single one. We refer to this work as the ET study (108).

The ET study modified this design in three ways: first, by relying on just one senior methodological expert rather than choosing a statistical reviewer from an expert list; second, by combining both interventions, with the senior methodological reviewer proposing specific changes based on relevant international

reporting guidelines; and, third, to avoid attrition by delaying the intervention until the decision had been made on whether or not to publish.

Figure 6 shows the scheme of the allocation of interventions for the ET study.

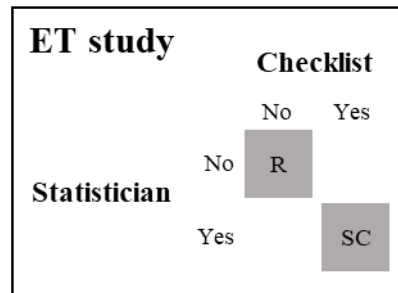


Figure 6. Scheme of the allocation of interventions for the ET study. *R=Reference; SC= both Statistician and Checklist.*

Figure 7 reproduces the flow chart of the ET study. From May 2008 to April 2009, 126 consecutive original papers were selected as possible study candidates. Of these papers, 34 (27%) were rejected after the conventional review, resulting in 92 randomized papers. The study designs of the included manuscripts were: 16 (17%) interventional studies, mainly before-after studies of only five randomized trials; 38 (41%) longitudinal studies; 26 (28%) cross sectional studies; and 12 (13%) studies of other types (mainly diagnostic studies). The authors ascertained protocol deviations in four papers in the conventional review group, which underwent an additional review based on reporting guidelines before the scheduled date. Again, they observed a modest increase in the MQAI scale measured at the end of the editorial process.

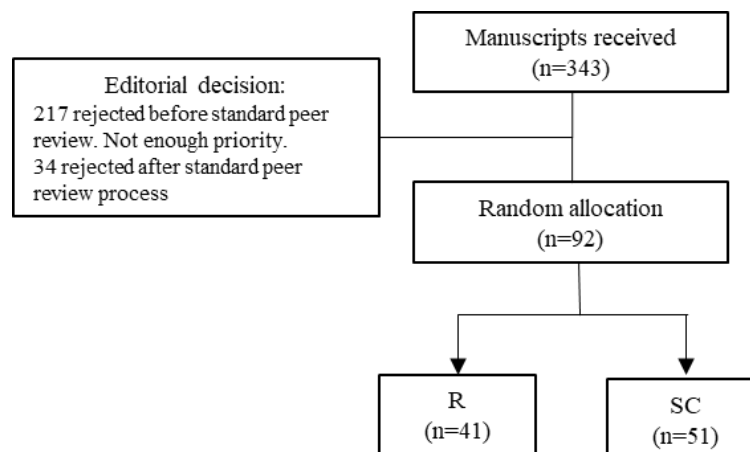


Figure 7. Summary of the flow chart of the ET study. *R=Reference; SC= both Statistician and Checklist.*

In addition to the standard editorial peer review revision (group R), both the ET and IQ studies had an intervention that added the revision of a statistician using RG (group SC). Therefore, we compare these groups for the main analysis. In the secondary analyses, we use the data of the IQ study to estimate separately the effect of the statistician (group S) and the effect of the publication guidelines (group C).

1.4. MAIN AND SECONDARY OBJECTIVES

In broad terms, the main objective of this thesis is to evaluate the long-term impact on biomedical research of interventions made during the editorial peer review process.

We work with all articles from two previous studies (ET and IQ studies) that applied interventions during editorial peer review in order to improve the quality of the manuscripts submitted to *Medicina Clínica* journal.

We do not have any references of previous works evaluating the long-term impact of published articles in biomedical research. However, we expect to have enough follow-up time, as the mean follow-ups are 13.35 (SD=0.31) for the IQ study and 9.29 (SD=0.31) for the ET study.

As we use the case from other studies (not designed to analyze the intervention effect on NC), we did not size our study to have enough power to perform a confirmatory study. Therefore, this is a proof of concept study, and we interpret the results as exploratory. Our intention is to provide the research community with the results of the present thesis in order to design a future confirmatory study.

This thesis presents different secondary goals according to its purpose. To define the distribution of NC is challenging because it could depend, among other things, on the time since article publication. Our secondary objectives are to find the count model that best fits NC and to become trained in a new statistical methodology based on recurrent events methods for estimating the intervention effect. Another secondary objective is to measure the separate effect of a statistical reviewer and a revision using reporting guidelines during the editorial peer review process.

Chapter 2

METHODOLOGY

2.1. DATA COLLECTION

From the data of the ET and IQ studies, we recover the article title and publication date. With this information and masked to the intervention group, I collected from WoS the number of citations and the publication date of each citation that the sample articles received. Later, we recovered the treatment group and baseline MQAI from the ET and IQ studies.

For the search in WoS, I included three items in the search tab:

- Publication name of the article, searched for topic to consider title changes (from submitted version to finally published);
- Medicina Clinica (Barcelona) as the journal;
- Publication year between 2004 to 2005 for IQ articles, and 2009 to 2010 for the ET articles.

Datasets for the analyses have different structures:

- Articles dataset: includes a row for article (identification variable (ID) of this dataset) with the common variables in Table 3. I use this dataset to analyze the number of citations at the end of the follow-up.
- Citations dataset: includes a row for each citation (ID of this dataset), adding an additional row for each article to represent articles without a citation. This longitudinal structure dataset has the same variables as the Articles dataset, but with the extra variables shown in Table 3. I use

this dataset to analyze the NC as recurrent events, taking into account the time when each citation appears.

	Variable	Description
Common variables	Article title	Article title (as finally published)
	Article ID	Unique identifier for each article
	Study	Abbreviation name of the study where the article belongs (ET or IQ)
	Group	Allocation treatment group (R or SC or S or C)
	Publication date	Publication date of the article
	Follow-up	Time from publication date to last NC update
	MQAI	Baseline MQAI obtained in the ET and IQ studies
	Number of citations	Number of citations of the article
Extra variables	Citation title	Citation title
	Citation publication date	Publication date of the article
	Time 1	Time from article publication to citation publication
	Time 2	Time between consecutive citations

Table 3. Description of the common variables for the Articles and Citations datasets, as well as for the extra variables belonging to the Citations dataset

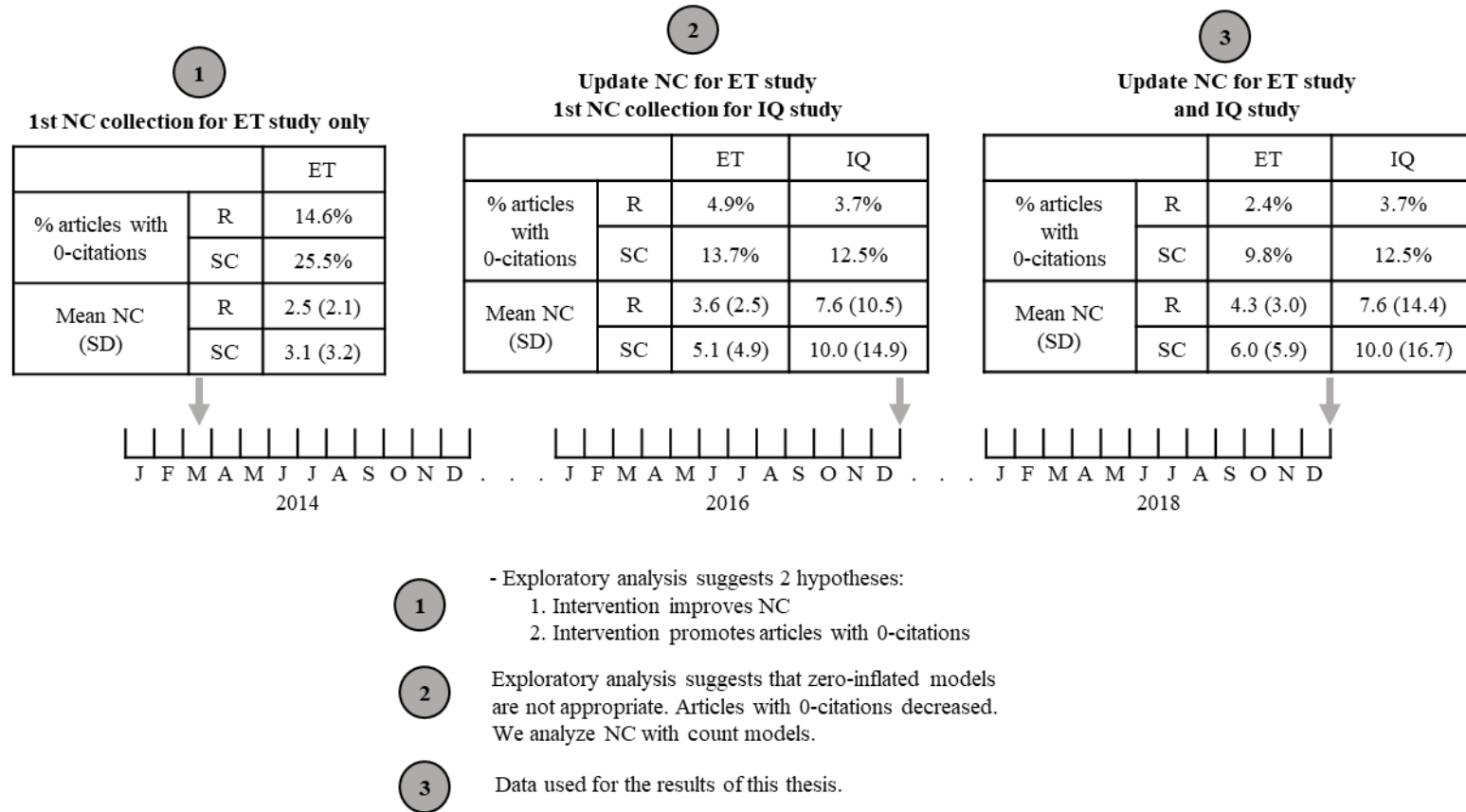


Figure 8. Scheme of the data collection evolution. *NC=Number of citations; R=Reference group; SC=Intervention group with the additional review of a statistician using reporting guidelines; SD=Standard deviation*

Since updating NC through WoS is relatively easy, I have collected it three times (see Figure 8):

- On 2014, the idea arose that it is possible to study further outcomes related to the long-term impact of the ET study, with the rationale that a better-reported paper should have better repercussions. The first time I collected NC was in March. I collected 5 years of follow-up data from the ET study. The objective of that analysis was to analyze the final Manuscript Quality Assessment Instrument (MQAI) scale (14) as a predictor of NC and the sum of IF, and to explore further effects of intervention on NC and IF. We found relationships between the final MQAI and both impact measures. The linear regression model slope for NC was 0.52 (95% CI: -0.04 to 1.08), and for the sum of IF it was 1.19 (95% CI: -0.17 to 2.55). These results are in Annex A.
This exploratory analysis suggested the hypothesis that intervention improves NC and promotes articles with zero citations. The mean (standard deviation or SD) was greater for the intervention group, with 3.1 (3.2) versus 2.5 (2.1) in the reference group. Furthermore, the proportion of zero citations was 14.6% for the reference group versus 25.5% for the intervention group. These characteristics lead us to think that zero-inflated models were appropriate for analyzing the data.
- Then, in December 2016, I updated the NC of the ET study and I collected the citations for the IQ study. An exploratory analysis suggested that zero-inflated models were not appropriate for a longer follow-up, since articles with zero citations decreased substantially (the ET study: R group from 14.6 to 4.9%, SC group from 25.5 to 13.7%). Annex C shows the results obtained with this update.
- Finally, the analyses detailed in this PhD thesis pertain to the data update until December 2018.

Three aspects guarantee the quality of the data collection process, the first is that, before collecting the final NC (2018), I trained in 2014 with the ET articles, which minimized data entry errors. The second is my blinding to the intervention group allocation. Finally, the third aspect is that a random 20 of the 191 articles were reviewed again by a second rater (Jordi Cortés), and no errors in the data collection process were detected.

WoS records citations in several databases apart from its core collection. At all collection times, I considered the NC included in all databases, ensuring consistency and comparability of the results.

2.2. COUNT DATA MODELS

This part of the thesis introduces the count models used to see the effects of the interventions over a count data outcome. Count data is discrete data with non-negative integer values that enumerate items, which in our case is the number of citations that an article receives during a given period.

First, we take a non-parametric approach, which assesses the intervention effects on NC without any model assumption. Then, we explain the most-used model for this type of data, the Poisson model.

After exploring the behavior of NC, we also analyze NC with other count models, such as the negative binomial, the zero-inflated Poisson and the zero-inflated negative binomial model. We use these analyses as sensitivity analyses to determine the robustness of the results by examining the methodology's degree of influence on them.

2.2.1. NON-PARAMETRIC EFFECT ESTIMATION

Assuming a predefined model to estimate an effect without knowing the distribution of the outcome can lead to erroneous estimates. Therefore, we consider calculating the point estimate of the effect and its uncertainty by means of a method without any modeling assumption, which we call a non-parametric methodology.

Count models like Poisson, NB, ZIP and ZINB estimate the average number of occurrences in a time interval, which in our case is the average number of citations per year. Therefore, we use the NC divided by the follow-up time as outcome for the non-parametric estimation. Thus, we can compare the results.

We use Jackknife (JK) resampling method, also known as “leave-one-out”, to estimate the bias and variability of an estimator $\hat{\phi}$ by using the values of $\hat{\phi}(Y)$ on subsamples Y_1, Y_2, \dots, Y_n . These subsamples have size $n-1$ and are assembled by successively excluding one observation from the full dataset with n observations, obtaining n estimates of $\hat{\phi}$.

The parameter we want to estimate is the mean; therefore, for $i=1, \dots, n$, we compute the mean $\hat{\phi}_{(i)}$ for each subsample that excludes the i -th data observation from the full dataset:

$$\hat{\phi}_{(i)} = \frac{1}{n-1} \sum_{j=1, j \neq i}^n y_j \quad [1]$$

Using the n JK estimates, $\hat{\phi}_{(1)}, \hat{\phi}_{(2)}, \dots, \hat{\phi}_{(n)}$, the standard error (SE) of the estimator is:

$$SE(\hat{\phi})_{JK} = \left\{ \frac{n-1}{n} \sum_{i=1}^n (\hat{\phi}_{(i)} - \hat{\phi}_{(\cdot)})^2 \right\}^{1/2}, \quad [2]$$

where $\hat{\phi}_{(\cdot)}$ is the average of the JK replicates:

$$\hat{\phi}_{(\cdot)} = \frac{1}{n} \sum_{i=1}^n \hat{\phi}_{(i)} \quad [3]$$

JK is a useful method for estimating the variability of some estimator, whereas this methodology does not expect awareness of the theoretical distribution of the estimator's standard error. However, this method is conservative since estimations of standard error are large, and JK performs inadequately when the estimator is not smooth enough (i.e. the median is a non-smooth statistic).

2.2.2. POISSON REGRESSION MODEL

The Poisson (log-linear) regression model is a generalized linear model (GLM) that is widely used for cross-sectional count data analyses. The GLM framework describes the dependence of a variable y_i ($i = 1, 2, \dots, n$) on a vector of regressors x_i .

To introduce the Poisson regression model I use the definition employed by Zeileis et al. (114) in the page 3 of their paper, with some changes on the names of the parameters to fit my thesis:

“The conditional distribution of $y_i|x_i$ is a linear exponential family with the probability density function

$$f(y|x, \delta, \psi) = \exp\left(\frac{y\delta - b(\delta)}{\psi} + c(y, \psi)\right), \quad [4]$$

where δ is the canonical parameter that depends on the regressors via a linear predictor, and ψ is a dispersion parameter that is often known. The functions $b(\cdot)$ and $c(\cdot)$ are known and determine which member of the family is used, e.g., the normal, binomial or Poisson distribution.

Conditional mean and variance of y_i are given by $E(y_i|x_i) = \mu_i = b'(\delta_i)$ and $\text{Var}(y_i|x_i) = \psi b''(\delta_i)$. Thus, up to a scale or dispersion parameter ψ , the distribution of y_i is determined by its mean.

The dependence of the conditional mean $E(y_i|x_i) = \mu_i$ on the regressors x_i is specified via

$$g(\mu_i) = x_i^T \beta, \quad [5]$$

where $g(\cdot)$ is a known link function and β is the vector of regression coefficients, typically estimated by maximum likelihood (ML) using the iterative weighted least squares algorithm.”

The straightforward distribution for modelling count data is the Poisson. The probability density function of the Poisson distribution is:

$$f(y|\lambda) = \frac{\lambda^y \exp(-\lambda)}{y!}, \quad [6]$$

where $y = 0, 1, 2, \dots$, represents the discrete random variable, such as the NC obtained by an article, with $\lambda > 0$. This probability function is of the form specified in [4]. The canonical link is $g(\lambda) = \ln(\lambda)$, resulting in a log-linear relationship between the mean and the linear predictor. The variance in the Poisson model is identical to the mean, thus the dispersion is fixed at $\psi = 1$.

Considering the canonical link, a possible regression equation for associating the p regression variables and the expected value of y is

$$\ln(\lambda) = \beta_0 + \sum_{i=1}^p \beta_i x_i \quad [7]$$

To model NC with and without adjusting for the baseline MQAI and including an “offset (t)”, the equations for our particular variables are:

$$\begin{aligned} \text{Non-adjusted model:} \quad & \log(\lambda) = \beta_0 + \beta_1 * \text{Group} + \text{offset}(t) \\ & = \beta_0 + \beta_1 * \text{Group} + \ln(t) \end{aligned}$$

$$\begin{aligned} \text{Adjusted model:} \quad & \ln(\lambda) = \beta_0 + \beta_1 * \text{Group} + \beta_2 * \text{baselineMQAI} + \text{offset}(t) \\ & = \beta_0 + \beta_1 * \text{Group} + \beta_2 * \text{baselineMQAI} + \ln(t) \end{aligned}$$

These can be rewritten as:

$$\text{Non-adjusted model: } \ln\left(\frac{\lambda}{t}\right) = \beta_0 + \beta_1 * \text{Group}$$

$$\text{Adjusted model: } \ln\left(\frac{\lambda}{t}\right) = \beta_0 + \beta_1 * \text{Group} + \beta_2 * \text{baselineMQAI}$$

Thus, this fits the mean-per-year NC.

The offset is a variable that denotes the exposure period (t in years) in the Poisson regression.

As we mentioned, the Poisson regression model must hold $Var(y) = E(y)$. When $Var(y) > E(y)$, the data presents overdispersion; and when $Var(y) < E(y)$, it presents underdispersion. Fitting the Poisson regression model when data do not hold the mean-variance equality assumption makes the estimations of standard errors inaccurate (115). If there is overdispersion, standard error estimates are biased downward (116).

If y_i are independent observations with corresponding values x_i of the predictor variables, then β can be estimated by maximum likelihood (ML). The ML estimator maximizes the log-likelihood function. The likelihood function for independent observations (117) is the product $\prod_i f(y_i|x_i, \beta)$ of the individual densities, conditioned on the regressors. The log-likelihood function is then the log of a product, which equals the sum of logs, or $\sum_i \ln f(y_i|x_i, \beta)$.

For the Poisson density [4], the log-density for the i th observation is

$$\ln(f(y_i|x_i, \beta)) = -\exp(x_i'\beta) + y_i x_i'\beta - \ln(y_i!) \quad [8]$$

So the Poisson ML estimator $\hat{\beta}$ maximizes

$$Q_N(\beta) = \frac{1}{N} \sum_{i=1}^N (-\exp(x_i'\beta) + y_i x_i'\beta - \ln(y_i!)) \quad [9]$$

The Poisson ML estimator is the solution to the first-order conditions

$$\frac{\partial Q_N(\beta)}{\partial \beta|_{\hat{\beta}}} = 0 \quad [10]$$

Or, equivalently

$$\frac{1}{N} \sum_{i=1}^N (y_i - \exp(x_i'\beta)) x_i|_{\hat{\beta}} = 0 \quad [11]$$

There is no explicit solution for $\hat{\beta}$ in [11]. Thus, ML estimates must be found by numerical methods. The second derivative of [11] is always negative for the Poisson regression, therefore this ML function

is always concave and gradient-based methods as Newton-Raphson are appropriate estimation techniques (117).

The Poisson model ML estimator is consistent if the specification of density for the data is correct.

Considering our main variable NC, three characteristics challenge this model:

- (1) It arises from independent and identical exponentially distributed inter-occurrence times between events, under the main assumption that events occur completely randomly over time. It is reasonable to think that the NC do not have the same probability of occurrence when an article already has one citation or more. After each new citation, the time to the next citation may decrease.
- (2) It assumes that the occurrences are independent. It would not be prudent to consider that citations are independent if each new citation gives the opportunity to cite the original work.
- (3) It assumes that there can be no more than one occurrence in each interval. Some articles may have citations within the same date.

R software provides a very flexible implementation of the general GLM framework in the function **glm()** (118) contained in the **stats** package. The family argument has to be specified “poisson” in the case of the Poisson model. The canonical link is $g(\lambda) = \ln(\lambda)$, resulting (as already mentioned) in a log-linear relationship between the mean and the linear predictor.

2.2.3. QUASI-POISSON AND NEGATIVE BINOMIAL REGRESSION MODEL

In many applications, the assumptions underlying the basic Poisson regression model are too restrictive (117). A common deviation from the basic Poisson model is that variance of the data exceeds the mean or overdispersion. The most common explanation for overdispersion is unobserved heterogeneity.

To deal with this overdispersion, it is usual to use a generalization of the Poisson regression model to include an extra parameter in the Poisson regression model which estimates how much larger the variance is than the mean. These models are the quasi-poisson (QP) regression model and the negative binomial (NB) regression model.

The QP regression model has two parameters that correspond to mean (λ) and overdispersion (θ), assuming that variance is a linear function of mean (119). Therefore, the first two moments for a random variable following a QP regression model ($y \sim P(\lambda, \theta)$) are $E(y) = \lambda$ and $Var(y) = \theta E(y) = \theta \lambda$.

This model uses quasi-likelihood (QL) function to estimate the coefficients. It has the advantage that the QL function does not need to specify the distribution of the observations, but only specify the mean-variance relationship (120).

Another alternative for modelling overdispersion is the negative binomial (NB) regression model. As in the QP model, the variance of a NB is a function of its mean and has an additional parameter for overdispersion. For a count random variable y following a NB, the variance of y is $Var(y) = \lambda + \lambda^2/\theta$.

The variance converges to the mean as the dispersion parameter gets larger, then the negative binomial turns into a Poisson distribution.

Equations for the NB regression model are the same as for the Poisson regression model, and it fits the mean-per-year NC. Furthermore, the NC's three characteristics described in the Poisson regression model section also challenge the NB regression model.

In summary, the QP and NB models have an advantage over the Poisson model in the flexibility of modeling the variance function, which can be appreciated in the case of heteroscedasticity.

The R software package **MASS** (121) provides the **glm.nb()** function for using ML to estimate the parameters of the NB model. The function **glm()** of the **stats** package with the family argument "quasipoisson" fits the QP model.

2.2.4. ZERO-INFLATED POISSON AND ZERO-INFLATED NEGATIVE BINOMIAL MODELS

Another deviation from the equidispersion assumption of the Poisson model is the frequently encountered feature that empirical data has a higher relative frequency of zero observations than any standard count regression model. A common approach to dealing with the excess zero problem is to use zero-inflated models (122) (123). With these models, the high frequency of zeros is accounted for, thus allowing extra probability mass at zero and reducing the probability mass for other frequencies.

Zero-inflated models assume that the data are a mixture of two separate data generation processes:

- Process 1 generates only zero counts. These zero counts are considered structural zeros. Authors usually model their probability with a logit model.
- Process 2 generates counts from either a Poisson or a NB model.

To determine which of the two processes generates an observation, the result of a Bernoulli trial is used.

For observation y , process 1 is chosen with probability ω , and Process 2 with probability $1 - \omega$,

$$y \sim \begin{cases} 0 & \text{with probability } \omega \\ g(y) & \text{with probability } 1 - \omega \end{cases} \quad [12]$$

where $g(y)$ is the probability function for Poisson or NB.

Thus, this mixture model has two independent sources of zeros:

- structural zeros above the expected zero frequency under Poisson or NB
- sampling zeros due to sampling variability under Poisson or NB

The probability of y is

$$P_{ZI}(y) = \begin{cases} \omega + (1 - \omega)g(y) & \text{if } y = 0 \\ (1 - \omega)g(y) & \text{if } y > 0 \end{cases} \quad [13]$$

For data with only an excess of zero counts, a Zero-Inflated Poisson model (ZIP) is proposed. However, for overdispersed count data with both extra zeros and unobserved heterogeneity, the Zero-Inflated Negative Binomial model (ZINB) is the appropriate one.

Mean and variance of ZIP (117) are

$$E(y) = \lambda(1 - \omega) \quad [14]$$

$$Var(y) = \lambda(1 - \omega)(1 + \lambda\omega) \quad [15]$$

And mean and variance of ZINB are

$$E(y) = \lambda(1 - \omega) \quad [16]$$

$$Var(y) = \lambda(1 - \omega)(1 + \lambda(\omega + \alpha)) \quad [17]$$

Both zero-inflated models present overdispersion: $Var(y) > E(y)$.

In R, zero-inflated count data models can be fitted with the `zeroinfl()` function from the `pscl` package (114). This function has a parameter to specify the data distribution (`dist`), where the option `dist="poisson"` is for ZIP and `dist="negbin"` is for ZINB.

2.2.5. COMPARING MODELS

Figure 9 shows the comparison of the abovementioned count models from an interpretability point of view. Poisson is the basic model, which estimates the effect of random citation (a group of articles with a random chance of having citations). When there is overdispersion (sample variance is greater than the mean), the Poisson model underestimates the standard error of the outcome. An excess of zero counts may cause this overdispersion, and that is when the ZIP model is pertinent. ZIP has the capability of differentially estimating the effect of random citation and the effect of non-citable articles (a group of articles that have a zero probability of being cited). When overdispersion is for a non-specific cause or for heterogeneity, then the NB model best fits the data by estimating an overdispersion parameter that allows the model to distinguish between random citations and this extra variability. Another model comes into play when there are both sources of overdispersion. ZINB model distinguish between random citation, effect of non-citable articles and overdispersion.

One way to compare the different models is to compare the sample distribution of the data to the average predicted distributions using the estimated models.

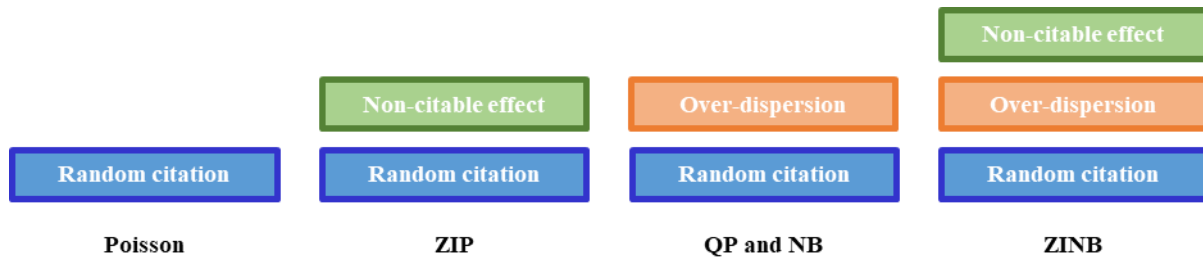


Figure 9. Comparison of the count models for the type of data to interpret

Kolmogorov-Smirnov statistic

The Kolmogorov-Smirnov statistic (K-S) quantifies the distance between two distributions. K-S is the maximum (“vertical”) distance between, for example, the empirical cumulative distribution function and the theoretical one. Its main advantage is that it is sensitive to differences in both the location and the shape of the cumulative distribution function.

For n independent and identically distributed (i.i.d.) ordered observations Y_i , the empirical distribution function F_n is defined as

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n I(Y_i), \quad [20]$$

where $I(Y_i)$ is equal to 1 if $Y_i \leq y$ and equal to 0 otherwise, behaving as an indicator function.

Then, for a given cumulative distribution function $F(Y)$ the K-S statistic is

$$D_n = \max\{|F_n(y) - F(Y)|\} \quad [21]$$

Thus, the closer the K-S statistic is to 0, the better the fit that the theoretical model has.

Akaike Information Criterion (AIC)

With the aim of comparing the fit between models, we take the information theory approach of the Akaike criterion (AIC). AIC estimates the relative amount of information lost by a given model. Then, the less information a model loses, the higher the quality of that model.

To calculate the AIC value of a model we use the expression

$$AIC = 2k - 2\ln(\hat{L}) \quad [22]$$

where k is the number of parameters in the model and \hat{L} is the maximum value of the likelihood function for the model.

Therefore, AIC includes a penalty for the number of estimated parameters discouraging overfitting and the lowest AIC value indicates a better fit.

Quasi models do not necessarily have a distributional form and do not depend on a likelihood. They are only characterized by their mean and variance. In 2002, Burnham and Anderson (124) developed quasi-AIC (qAIC) to compare within quasi models but the validity of this approach for comparing with AIC has not been demonstrated (120).

Note that AIC and qAIC say nothing about the absolute quality of a model, only the quality relative to other models.

2.3. JOINT EFFECT ESTIMATION

The meta-analysis (MA) methodology is the statistical procedure for combining data from multiple studies. MA identifies a common effect when the treatment effect (“effect size”) is consistent from one study to the next. Then, we use this method to estimate the joint effects in the ET and IQ studies for the common intervention groups (R versus SC).

A MA is the synthesis of K (number of studies in the MA) compatible effects (γ_j) giving greater weight to studies with less uncertainty (V_j), that is, more precision ($w_j = 1/V_j$), for $j = 1, \dots, K$.

There are two main approaches to applying a meta-analysis methodology: the fixed-effects model and the random-effects model. A fixed-effects MA usually takes the inverse variances of the estimates as weights, and interpretation relies on the assumption of a common effect underlying every study. A random-effects MA incorporates the underlying among-study variation of effects into the weights.

For both the fixed-effects and the mixed-effects models, the parameter estimate is given by

$$\hat{\gamma} = \frac{1}{\sum_{j=1}^K w_j} \sum_{j=1}^K w_j \hat{\gamma}_j \quad [25]$$

and

$$\text{Var}(\hat{\gamma}) = \frac{1}{(\sum_{j=1}^K w_j)^2} \sum_{j=1}^K w_j^2 \sigma_{\gamma_j}^2 \quad [26]$$

The difference between the fixed-effects and random-effects models is in the weight w_j . In the fixed-effects model, $w_j = 1/\sigma_{\delta_j}^2$, where $\sigma_{\delta_j}^2$ is the within-study variance. In the random-effects model, $w_j = 1/(\sigma_{\delta_j}^2 + \sigma^2)$, where σ^2 is between-study variance.

The fixed-effects model assumes that the differences between the effect sizes of individual studies are due only to the fact that the studies use different samples of subjects (125). That is, the variability is exclusively a result of random sampling errors.

To estimate the joint effect, the **rmeta** package in R provides the **meta.summaries()** function.

2.4. RECURRENT EVENTS METHODS

Over the last few decades, many powerful survival methods have been developed for recurrent event data by extending Cox's proportional hazard regression. These methods can be differentiated into two groups by the way they deal with within-subject correlation: variance-corrected models and frailty models. Most popular variance-corrected models are Anderson-Gill (126), Prentice Williams and Peterson (127), and Wei Lin and Weissfeld (128). Contrary to variance-corrected models, frailty models assume that correlation among recurrent events is due to some subjects (articles) being more prone to events (citations) as compared to others because of some unobserved/unknown factors.

Frailty models directly incorporate into the model estimation a frailty term, which follows a specific distribution. The reasons for this are to accommodate the unobserved factors and to correct the estimation of regression parameters. Variance corrected models adjust the variance-covariance matrix in order to deal with the within-subject correlation.

Considering the properties of variance-corrected models and frailty models (see Annex F), the latter are more appropriate for the analysis of NC. Frailty models capture within-article correlation due to citation dependence and heterogeneity.

Using recurrent events methods, we try to see if the main intervention effect estimation remains similar, independently of the statistical method, thus giving robustness to the results. Another aspect we want to explore is if these techniques, which consider more information, lead to narrower 95% CI.

Apart from frailty models, we find the mean cumulative function (MCF) useful for visualizing the tendency of the NC over the two intervention groups. The shape of MCF can give us an idea of how long we should wait to study the long-term effect of the intervention on the NC of articles in a medium-impact journal like *Medicina Clínica* (Elsevier, Barcelona).

2.4.1. MEAN CUMULATIVE FUNCTION

Recurrent events data analysis describes every unit in the population by means of a cumulative history function for the number of recurrences. In other words, it represents the accumulated number of events over time. The average of all population curves passing through the vertical line at each time t is known as the MCF.

The MCF method assumes that:

- (i) The sample units represent a random sample of the target population.
- (ii) The cumulative history functions of all sample units are statistically independent of their censoring times (i.e., random censoring, also called non-informative censoring) (129).

Nelson (130) describes a simple nonparametric estimator of the MCF, which is analogous to the Nelson-Aalen estimator for the cumulative hazard function of lifetime data (131), as explained below.

Formally, let $N_i(t_j)$ denote the number of citations happening to article i at time point t_j , where $1 \leq i \leq n$ articles and $1 \leq j \leq m$ time points. Furthermore, each article has a censoring time ζ_i , after which the citations are not known. Let $O_i(t)$ denote whether article i is observable at time t , meaning that $O_i(t) = I[t \leq \tau_i]$. The $N_i(t_j)$ is defined as zero when there is no event happening to article i at time t_j or when the article is censored before any citation, i.e., $O_i(t_j) = 0$. Finally, we denote by $NT_i(t) = \sum_{j:t_j \leq t} N_i(t_j)$ the total number of citations accumulated by article i in the interval $(0, t_j]$.

To define the aggregate values over all articles, denote by $N_o(t) = \sum_{i=1}^n N_i(t)$ the total number of citations at time t and by $O_o(t) = \sum_{i=1}^n O_i(t)$ the total number of articles observable at time t . The MCF represents the population mean number of recurrent events by certain times. $MCF(t) = E[NT(t)]$, where $NT(t)$ is a random variable for the number of events that have occurred up to time t . The MCF estimates the expected value $E[NT(t)]$ as a cumulative sum of the total NC over all event times divided by the number of articles observed at time t :

$$E[NT(t)] = \sum_{j:t_j \leq t} \frac{N_o(t_j)}{O_o(t_j)} \quad [27]$$

A robust variance estimate is (132):

$$\text{Var}[E[NT(t)]] = \sum_{i=1}^m \left(\sum_{j:t_j \leq t} \frac{O_i(t_j)}{O_o(t_j)} \left(N_i(t_j) - \frac{N_o(t_j)}{O_o(t_j)} \right) \right)^2 \quad [28]$$

Any estimate based on a sample has uncertainty associated with its sampling. Thus, it is important to estimate confidence intervals that quantify the probable range of the MCF at each given time point t :

$$95\% \text{ CI} = E[NT(t)] \pm 1.96\sqrt{\text{Var}[E[NT(t)]]} \quad [29]$$

To assess the intervention effect, we compare the difference of the MCFs of the R and SC groups, denoted by $E[NT_{SC}(t)]$ and $E[NT_R(t)]$. The estimate of each citation time point MCF difference is

$$E[\Delta NT(t)] = E[NT_{SC}(t)] - E[NT_R(t)] \quad [30]$$

Since the sampling of articles is independent due to randomization, the variance of the difference is the sum of the MCF variances that were estimated using the previous formula [28]:

$$\text{Var}[E[\Delta NT(t)]] = \text{Var}[E[NT_{SC}(t)]] + \text{Var}[E[NT_R(t)]] \quad [31]$$

The estimate of the difference $E[\Delta NT(t)]$ between MCFs, with corresponding CIs based on $\text{Var}[E[\Delta NT(t)]]$, can be plotted similarly to the MCFs. CIs based on the MCF show the expected difference and the probable region of this difference for every time point over a long duration

R software has the *reda* package to analyze the recurrent event data by means of the MCF methodology. It implements the `mcf()` function to estimate the MCF of the data and the `mcfDiff()` function to estimate the difference between two MCFs at each time.

2.4.2. FRAILTY MODELS

Frailty models provide a suitable way to introduce random effects into a survival model to account for within-subject correlation and unobserved heterogeneity across subjects, which in our case are the articles. These models assume that correlation among recurrent events over time is due to some individuals being more susceptible to having a recurrent event than others because of some unobserved or unknown factor (133).

Thus, a frailty model is a random effects model for time-to-event data, where the random effect (the frailty) has a multiplicative effect on the baseline hazard function. Frailty models are an extension of Cox's model for introducing frailty terms (nonobservable random covariates).

The frailty term follows a specific distribution, but there are no guidelines on how to select an appropriate frailty distribution for a given scenario. The gamma distribution is the most commonly used in the estimation of this frailty term, although there are other distributions for frailty estimation.

Some authors have distinguished (133) between two categories of frailty models: univariate frailty models, which consider univariate survival times; and multivariate frailty models, which account for multivariate survival times (e.g., competing risks, recurrence of events in the same individual, and occurrence of a disease in relatives).

We are interested in multivariate frailty models, as they account for the presence of dependence between event times. To account for dependence, the frailty model introduces a random effect called frailty that is denoted by Z . This frailty term describes the common risk within each article acting as a multiplicative factor in the hazard function.

The definition of a frailty model in terms of conditional risk is:

$$h(t_{is}|Z_i) = Z_i h(t_{is}), \quad [32]$$

where the frailty term Z_i is added to the classical Cox model. The term Z_i is a nonobservable random variable that varies over the sample and increases the individual risk if $Z_i > 1$, and it decreases if $Z_i < 1$. When interpreting the model, the hazard ratios (HR) of the covariate x are the same as they are for the Cox model without frailty.

When this frailty term is cluster-specific, as in recurrent events, the frailty model is known as a shared frailty model.

Under the shared frailty model, citations of article i are supposed to share the same frailty (134). The conditional hazard function at time t of citation s in article i is:

$$h_{is}(t|x, \beta, Z) = Z_i h_{is}(t) = h_0(t) e^{x_{is} \beta Z_i}, \quad [33]$$

where h_{is} is the conditional hazard function for the s^{th} citation from the i^{th} article (conditional on Z_i); $h_0(t)$ is the common baseline hazard function; β is the regression coefficient vector of dimension p ; x_{is} represents the vector of covariates of citation s in article i ; Z_i is the random effect on the i^{th} article. The frailty terms Z_1, Z_2, \dots, Z_n are independent and identically distributed with known density $f(\cdot | \theta)$ and unknown parameter θ .

As in the proportional hazards model, parametric or non-parametric forms of baseline hazard can be assumed in frailty models (135). The parametric model performs the parameter estimation by means of the maximum marginal likelihood approach (136). The semi-parametric model leaves the baseline risk function unspecified, since the shape of the baseline hazard function is often unknown in practice. For parameter estimation, the semi-parametric model uses penalized partial likelihood, expectation maximization or hierarchical-likelihood methods.

The frailty term Z_i accounts for the extra variability from unobserved risk that are not accounted for in the model. If we assume that Z_i follows some distribution, then we can estimate the frailty model. The goal is to estimate the frailty variance, as it is used to determine the degree of heterogeneity in the study

population. When the variance is small, the values of frailty are around one. When the variance is large, there is dispersion on the values of frailty, making that the frailty promotes greater heterogeneity into the individual hazards.

The shared frailty model has some limitations, as it assumes that the unobserved covariates are the same within the cluster, or that the dependence parameter and the population heterogeneity are confounded.

The choice of the frailty distribution is very important, as the shape of the distribution plays an important role in frailty models. The tails of the distributions can determine the type of dependence. Distributions with a large right tail such as positive stable distributions lead to strong early dependence, whereas distributions with a large left tail such as gamma lead to strong late dependence (137).

The one-parameter gamma distribution is the most widely used frailty distribution (138), since it is very tractable. We will focus on the gamma frailty model, although we will compare the results with other frailty distributions to see the robustness of the findings.

Gamma frailty model

Suppose a random variable $z > 0$ is gamma distributed with shape parameter $\alpha > 0$ and scale parameter $\theta > 0$. The probability density function of z is:

$$f_z(t) = \frac{\theta^\alpha t^{\alpha-1} \exp(-\theta t)}{\Gamma(\alpha)}, \quad t > 0, \quad [34]$$

where

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} \exp(-s) ds \quad [34]$$

is the gamma function.

The expected value and variance of the gamma distribution are

$$E(z) = \frac{\alpha}{\theta} \quad \text{Var}(z) = \frac{\alpha}{\theta^2} \quad [35]$$

The gamma distribution $Gamma(\alpha, \theta)$ takes a variety of shapes as α varies. When $\alpha = 1$, it is identical to the exponential distribution. The relationship between α and the hazard function is

$$h(t) = \begin{cases} \text{decreasing} & \text{when } 0 < \alpha < 1 \\ \text{constant} & \text{when } \alpha = 1 \\ \text{increasing} & \text{when } \alpha > 1 \end{cases}$$

In gamma frailty models, the restriction $\alpha = 1$ is used, which results in an expectation of 1. The variance of the frailty variable is then $1/\theta$. Assuming that the frailty term z is distributed as gamma with $E(Z) = 1$ and $Var(Z) = \theta$, then the distribution function of the frailty term z is the one-parameter gamma distribution:

$$f(z) = \frac{z^{\frac{1}{\theta}-1} \exp(-z/\theta)}{\Gamma(1/\theta)\theta^{1/\theta}}, \quad \theta > 0 \quad [36]$$

where $z > 1$ indicates that article i has a higher probability of being cited, whereas $z < 1$ indicates that article i has a lower probability of being cited.

The maximum likelihood estimation in the gamma frailty model is straightforward, since we can easily integrate the frailties into the likelihood function and obtain the parameter estimates using classical maximum likelihood techniques.

Shared frailty models with R software

There are diverse estimation approaches available in the R software for fitting a shared frailty model (135), as shown in Table 4.

The baseline hazard function in a parametric model has known parametric distribution with unknown parameters to estimate. The method to estimate these parameters is the maximum marginal likelihood (MML) approach (136). The `parfm` package (139) provides with several parametric frailty models.

Generally, the distribution of the baseline hazard function is unknown, therefore it is advisable to use a semi-parametric model that leave the baseline hazard function unspecified. Under this semi-parametric framework, `survival` (140) and `coxme` (141) packages employ the penalized partial likelihood (PPL). The `frailtypack` package obtains the parameter estimates by nonlinear least squares (NLS) methodology.

To test the performance of each R package, we simulate data with an exponential baseline hazard time and frailty term following a gamma distribution – the most usual frailty distribution assumption when fitting shared frailty models. We compare the bias of the effect estimation to decide which package to use for estimating our real data with the formula:

$$Bias(\beta) = \frac{\sum_{r=1}^{nsim} \hat{\beta}_r}{nsim} - \beta, \quad [37]$$

where $nsim$ is the number of simulations, $\hat{\beta}_r$ are the estimates of the effect within each of the $nsim$ simulations, and β is the assumed true value of the intervention effect. Annex G describes in detail the simulation study.

Package::function	Baseline hazard function	Estimation procedure	Frailty distribution
parfm::parfm	Parametric	MML	Gamma, PS, IG
survival::coxph	Nonparametric	PPL	Gamma, LN
coxme::coxme	Nonparametric	PPL	LN
frailtypack::frailtyPenal	Nonparametric	NLS	Gamma, LN
frailtyEM::emfrail	Nonparametric	ML	Gamma, PS, IG, PVF

Table 4. R packages and functions for fitting shared frailty models for recurrent events. *MML= maximum marginal likelihood; PPL=penalized partial likelihood; NLS= nonlinear least squares; ML: maximum likelihood; PS=positive stable; IG=inverse Gaussian; LN=log-normal; PVF=power variance function*

2.5. REPRODUCIBILITY (Shiny application)

The scientific method relies on reproducibility, the benchmark for the reliability of an experiment.

There is a growing interest in evaluating and ensuring the transparency and reproducibility of the published scientific literature. Since 2014, there have been intensified efforts to promote open science practices across the biomedical literature (142). Maintaining reproducibility requires that all data management, analysis, and visualization steps behind the results presented in a paper are documented and available in full detail. Reproducibility here means that someone else should be able to obtain the same results, given all the documented inputs and the published instructions for processing them.

It is almost impossible to reproduce the results of a study with only the manuscript of the publication; one also needs the code and the raw data used. Furthermore, to ensure reproducibility, the best strategy is to provide executable code and data (Figure 10).

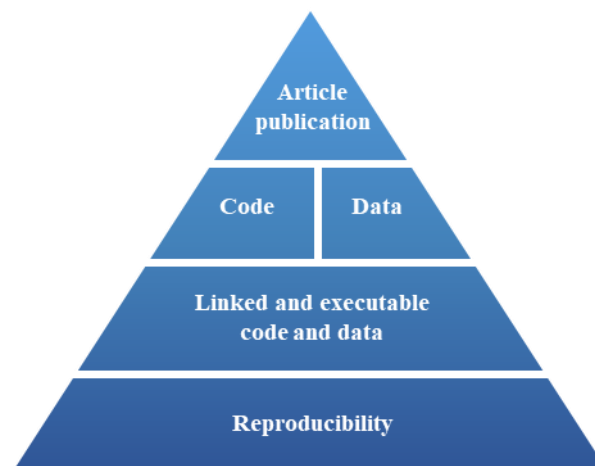


Figure 10. Steps towards reproducibility

With the aim of providing reproducibility for this research, I have created an application with the Shiny tool, which allows fully replicating the results obtained. This application provides the data and the obtained results in an easy and interactive interface.

The Shiny tool is a framework for building interactive web applications (apps) using the R language. It allows users to interact and explore data, and to understand models and its dependency on different parameters without managing the code.

The link to the application is: <http://shiny-eio.upc.edu/pubs/NumberCitations/>

It has two tabs. The first tab consists of the dataset table used to perform the analyses. Table 5 shows the name of the dataset table variables and their meanings. Through the application menu, it is possible to see all the data and filter the data separately, which in this example are of the ET and IQ studies. It also allows choosing the number of registers to display and provides the option of downloading the data in order to work with it.

Name of the variable	Explanation
Title	Title of the original randomized article in the ET and IQ studies
Study	Study to which it belongs (ET or IQ)
Study design	Type of study design
Group	To which intervention group it was randomized (R; S; C; SC)
Publication date	Date of publication of the article in the Medicina Clínica journal
Follow-up	Time of follow-up to 31/12/2016
Goodman scale	The baseline MQAI scale value evaluated in the ET and IQ studies
NC	Number of citations collected using WoS that the article has received at the end of the follow-up

Table 5. Names of the variables and their meanings that appear in the first tab of the shiny app.

The other tab shows the results obtained using graphics. These are the results of the main comparison (R vs. SC) for both studies (ET and IQ) and the joint effect estimation. It is possible to see the results obtained by the Poisson model and nonparametric methodology. You can choose to view the analyses according to type of study. In the case of estimating with the Poisson model, you can choose how to estimate the standard error through the Jackknife or Wald methods. It also shows the results of secondary comparisons: the effect of the statistical reviewer (R + C vs. S + SC) and the effect of the publication guidelines (R + S vs. C + SC). For each display, there is a small explanation of the result shown.

Chapter 3

RESULTS

3.1. DESCRIPTIVE RESULTS

The IQ study includes 99 articles published between 4 February 2005 and 12 May 2006, with a mean (SD) follow-up period of 13.35 (0.31) years. These publications received 1037 citations (mean 10.47, SD 15.94). The main analysis includes only the 24 and 27 papers allocated to the SC and R groups, which received 545 citations (mean 10.69, SD 15.37).

The ET study includes 92 articles published between 24 June 2009 and 3 April 2010, with a mean (SD) follow-up period of 9.29 (0.31) years. 41 and 51 papers belong to the R and SC groups, respectively. They received 484 citations (mean 5.26, SD 4.86).

Table 6 shows the descriptive results for each study and intervention group of the NC collected until December 2018.

For the IQ study, Table 6 uses a shaded style to differentiate the descriptive data of groups not included in the main analysis, i.e., the S and C groups. The R and SC groups have obtained 268 and 277 citations, with means (variance) of 9.93 (207.30) and 11.54 (277.48), respectively. The mean (SD) annual rate is 0.74 (1.08) for the R group and 0.86 (1.22) for the SC group. In the R group there is 1 (3.70%) article with 0 citations, and the SC group has 3 (12.50%).

During the 9.29 years of follow-up period for the ET study, the R and SC groups obtained a total of 176 and 308 citations, respectively. The R group has a mean (variance) of 4.29 (9.11) citations and the SC group's is 6.04 (34.24) citations. The mean (SD) annual rate is 0.46 (0.32) for the R group and 0.66

(0.66) for the SC group. In the R group, there are 1 (2.44%) articles without any citations, whereas the SC group has 5 (9.80%) articles. The proportion of articles with 10 or more citations is higher in the SC group, with 11 (21.57%) versus 2 (4.88%) articles in the R group.

In both studies, the intervention groups have higher means and standard deviations of NC, higher annual rates, and higher proportions of articles with 0 citations (Figure 11).

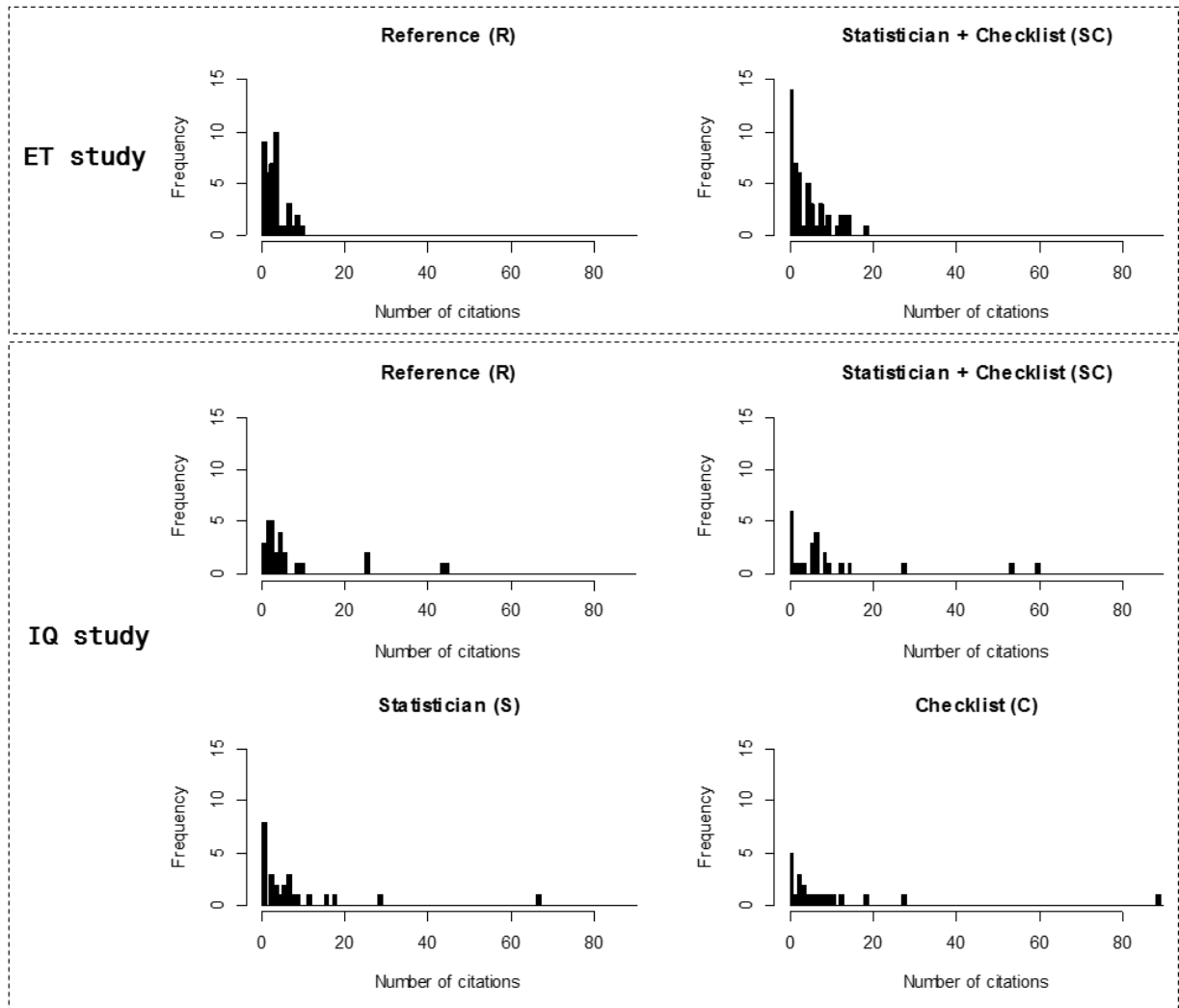


Figure 11. Number of citations by study and intervention group

Figure 12 displays the data in a recurrent events perspective for the ET study. It shows separate event plots for each article, where crosses indicate the days that citations occurred. In cases where there were more than one citation for a given article on a given day, I slightly separate the cross. This figure provides an impression of the frequency and patterns of citations throughout the follow-up time for each article.

		N	Number of citations				Sum NC	Articles with 0 citations	Annual rate
			Mean	Variance	Median	IQR		N (%)	Mean (SD)
IQ study	R	27	9.93	207.30	5.00	8.00	268	1 (3.70%)	0.74 (1.08)
	S	26	9.46	202.98	5.00	6.50	246	4 (15.38%)	0.71 (1.07)
	C	22	11.18	379.68	5.50	9.25	246	2 (9.09%)	0.83 (1.42)
	SC	24	11.54	277.48	7.00	8.50	277	3 (12.50%)	0.86 (1.22)
	Total	99	10.47	254.01	5.00	8.00	1.037	10 (10.10%)	0.78 (1.18)
ET study	R	41	4.29	9.11	4.00	3.00	176	1 (2.44%)	0.46 (0.32)
	SC	51	6.04	34.24	4.00	7.00	308	5 (9.80%)	0.66 (0.66)
	Total	92	5.26	23.58	4.00	5.00	484	6 (6.52%)	0.57 (0.54)

Table 6. Number of citations by study and intervention group until December 2018. Groups not included in the main analysis are in a shaded style.

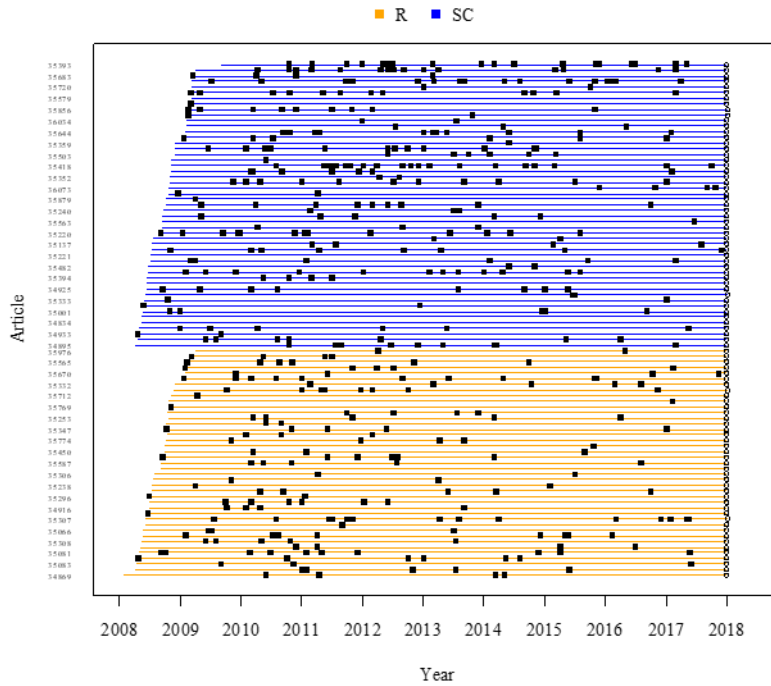


Figure 12. Evolution of citations by group in the ET study

To visualize the increment in citations over time, we plot the MCF for articles in the ET study (Figure 13). It shows that the growth is not constant over the follow-up. During the first year after publication, articles receive a moderate increment in NC. The most important increment in citations is between the first and fourth year after publication. After the fourth year, the increment starts to lose strength.

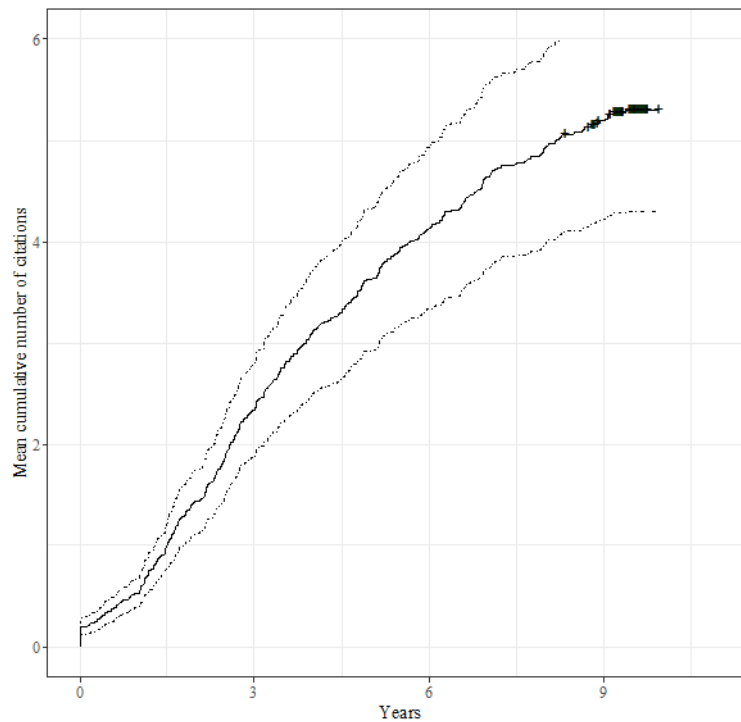


Figure 13. Mean cumulative function for the ET study articles

3.2. EFFECT OF A STATISTICAL REVIEWER USING REPORTING GUIDELINES

Figure 14 displays the adjusted and non-adjusted estimation of the SC additional intervention effect using the Poisson model. This figure shows narrow 95% CI when using the Wald standard errors, or what is the same standard errors from the information matrix. These 95% CI do not seem trustworthy since it compares small sample sizes. This may indicate a violation of the distribution assumption that the variance equals the mean. In such cases it is recommended (117) using robust standard errors for the parameter estimates, as resampling methods.

NC have overdispersion and the Poisson model is not an optimal model to fit the data (see Annex B).

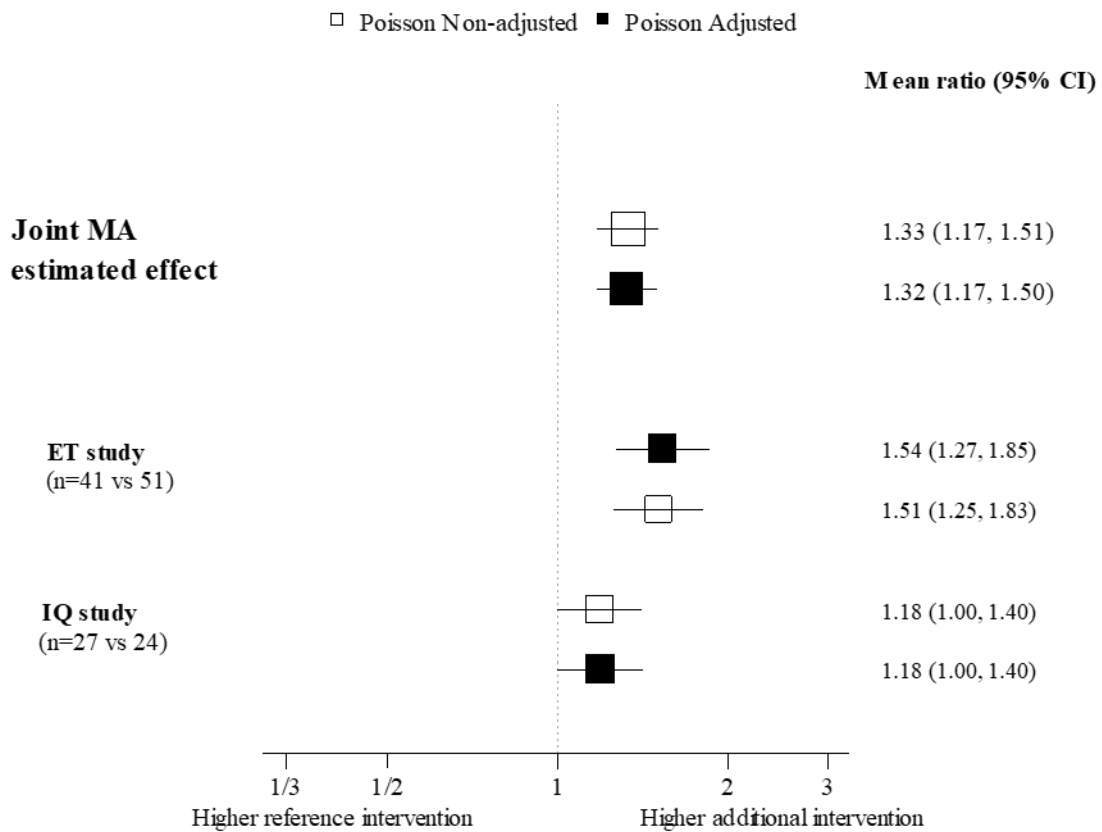


Figure 14. Citations-per-year mean ratio for the main comparison (R vs. SC) using the Poisson model with Wald standard errors

Figure 15 shows three different estimates of the SC additional intervention using the mean citations-per-year ratio:

1. Raw estimate: without using any model assumption, I use the JK resampling method with the outcome NC/Follow-up for each article
2. Poisson non-adjusted: this estimate assumes a Poisson regression model with group and offset(follow-up) covariates
3. Poisson adjusted: this estimate assumes a Poisson regression model with group, baseline MQAI value and offset(follow-up) as covariates

All reported 95% CI come from the JK method.

For the ET study, the intervention increases the mean citations-per-year by 44% (95% CI: 1 to 107%) using raw estimation, by 54% (95% CI: 4 to 127%) using Poisson non-adjusted and by 51% (95% CI: 3 to 123%) using Poisson adjusted.

For the IQ study, the intervention increases the mean citations-per-year by 16% (95% CI: -52 to 180%) using raw estimation, by 18% (95% CI: -51 to 183%) using Poisson non-adjusted and by 18% (95% CI: -51 to 184%) using Poisson adjusted.

For both the ET and IQ studies, point estimates of the intervention are positive. CIs for the IQ study are wider. One possible explanation for this is the number of articles under comparison (R=27 vs. SC=24) compared with the ET study, which has more articles per group (R=41 vs. SC=51).

The joint MA estimated effect is 40% (95% CI: 1 to 94%) for the raw estimation, 47% (95% CI: 4 to 109%) for Poisson non-adjusted and 45% (95% CI: 2 to 105%) for Poisson adjusted. MA effects are similar to those of the ET study, as they have more weight in MA.

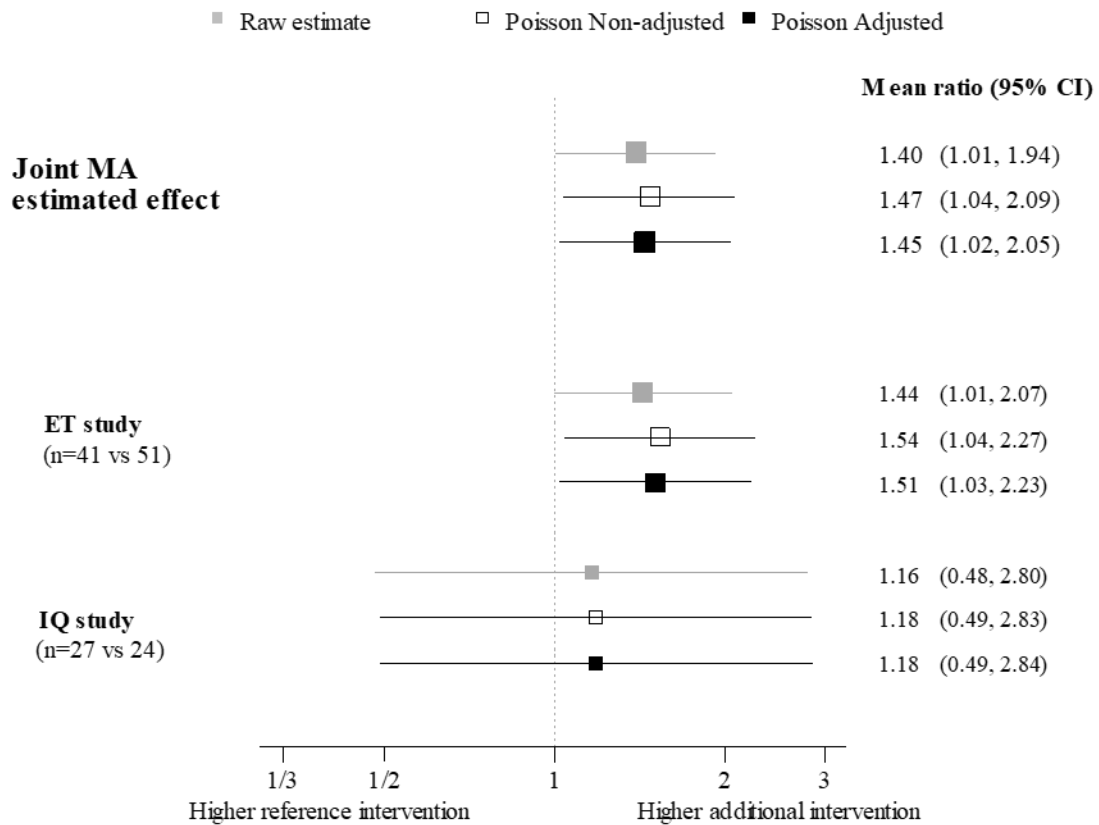


Figure 15. Citations-per-year mean ratio for the main comparison (R vs. SC)

Figure 16 shows the estimated mean-per-year ratio with the QP, NB, ZIP and ZINB models.

Point estimated effects of the mean-per-year ratio are greater than 1 for all models in both the ET and IQ studies, as in the MA analysis. Furthermore, the joint MA estimated 95% CI that are greater than 1 for all models and adjustments. The NB model fits the NC better than the other models (Figure 17, Table 7, Table 8 and Table 9). The joint MA estimated effect is 61% (95% CI: 8 to 138%) for the non-adjusted and 59% (95% CI: 5 to 140%) for the adjusted NB model.

Therefore, the results of the citation-per-year mean ratio for the Poisson regression model and nonparametric estimations are consistent with the results obtained from the other count models, all of which show that NC is improved by the additional intervention of including a statistical reviewer that uses reporting guidelines in the editorial process.

Figure D.1 of Annex D shows the odds ratio (OR) of non-citable articles (extra zeros) estimated by ZIP and ZINB. Extra zero estimates have wide intervals, indicating inaccurate estimations, which is most probably due to the absence of high extra zero counts for the NC collected until 2018. Considering the

MA of the ZIP non-adjusted model, the estimated odds of observing an excess zero in the intervention group is 11.44 times (95% CI: 0.00 to 6.17×10^{10}) the odds in the reference group. This wide CI indicates that zero-inflated models are not appropriate for our data, accordingly with Table 7 and Annex E.

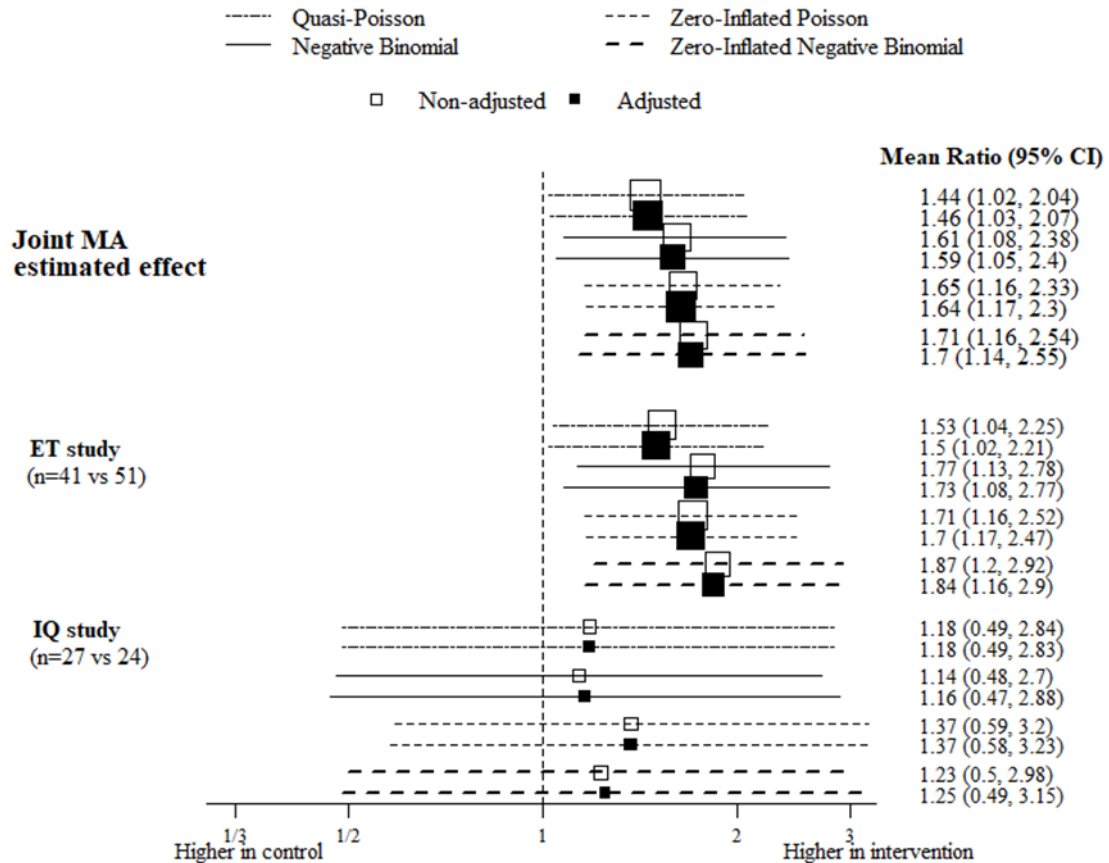


Figure 16. Citation-per-year mean ratio using QP, NB, ZIP and ZINB models

Figure 17 shows the empirical and theoretical cumulative distribution functions (CDFs) of the sample and the different models tested, respectively. It shows that the model that best fits the data is the NB. Figure E.1 of Annex E shows the histogram of the ET study data with theoretical densities of Poisson, QP, ZIP, NB and ZINB, thus supporting the use of NB to fit the NC data.

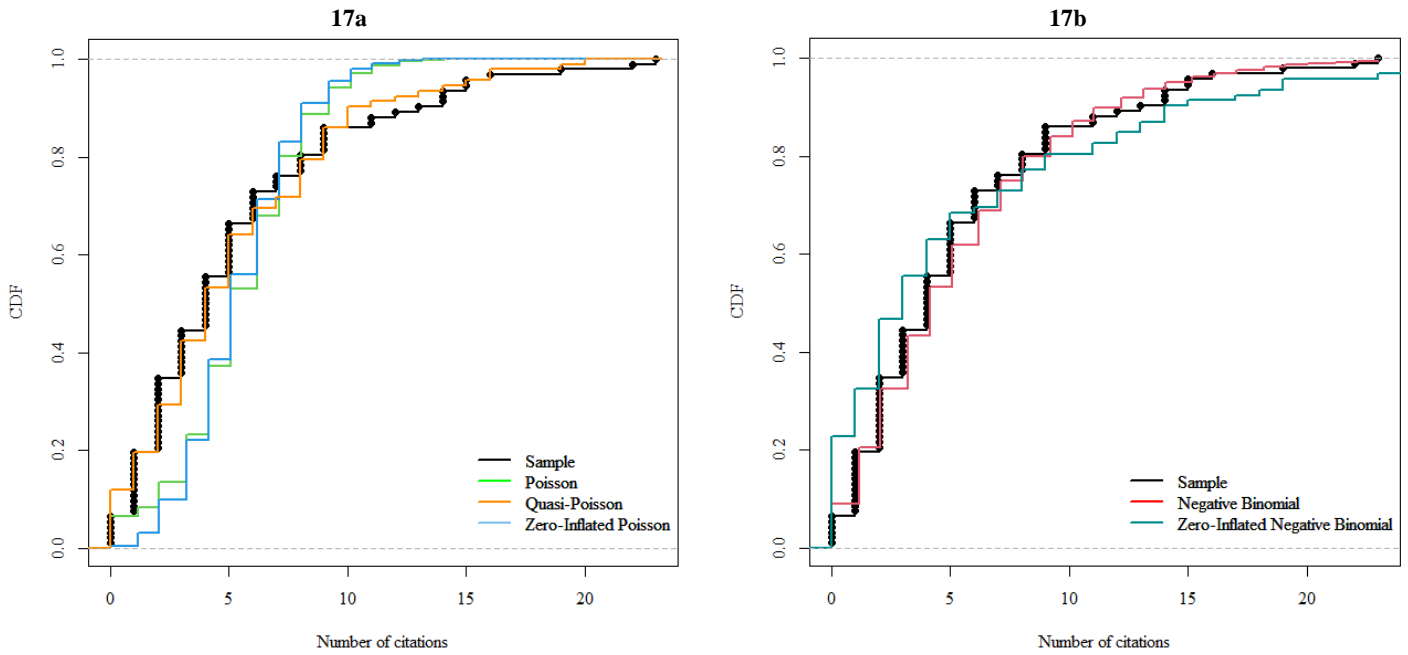


Figure 17. Empirical and theoretical CDFs for the ET study. For clarity, this figure shows two plots: 17a for the CDFs of Poisson, QP and ZIP, and 17b for NB and ZINB

The Kolmogorov-Smirnov statistic (K-S) in Table 7 shows the absolute maximum distance between the CDF of the sample and the CDFs of the other models considered. The closer K-S is to 0, the better the fit. For the ET and IQ studies, the model that best fits the NC is the NB model. The maximum K-S distance between the NB model’s CDF and the sample CDF is 0.08 (95% CI: 0.04 to 0.16) for ET study and 0.16 (95% CI: 0.08 to 0.27) for IQ study.

	ET study		IQ study	
	K-S	95% CI	K-S	95% CI
Poisson	0.25	(0.18, 0.30)	0.49	(0.41, 0.55)
Quasi-Poisson	0.09	(0.04, 0.15)	0.18	(0.10, 0.29)
Negative Binomial	0.08	(0.04, 0.16)	0.16	(0.08, 0.27)
Zero-Inflated Poisson	0.22	(0.15, 0.29)	0.45	(0.35, 0.53)
Zero-Inflated Negative Binomial	0.59	(0.49, 0.67)	0.75	(0.65, 0.84)

Table 7. Kolmogorov-Smirnov statistic by study and count model

After estimate the effect of a statistical reviewer using reporting guidelines by the different count models, we note consistency of the effect independently of the chosen model. Table 8 and Table 9 show a summary of the estimated effects and AIC values for the different models. In both ET and IQ studies the minimum AIC corresponds to the NB model, which agree with the shortest KS distance (Table 7).

Since the QP model is not calculated based on the likelihood approach, it does not have AIC estimation. It is possible to compute qAIC but it is not comparable with AIC, therefore it is not offered. Other model selection criterion such as the Bayesian information criterion (BIC) relies also on the likelihood approach, therefore quasi BIC is not neither comparable to BIC.

Raw estimation of the effect is in the line with the estimation of the count models. It reaffirms that the effect of the intervention is free of model selection bias.

ET study	Non-adjusted by baseline MQAI		Adjusted by baseline MQAI	
	Mean ratio (95% CI)	AIC	Mean ratio (95% CI)	AIC
Poisson	1.54 (1.04, 2.27)	754.34	1.51 (1.03, 2.23)	747.47
Quasi-Poisson	1.53 (1.04, 2.25)	-	1.50 (1.02, 2.21)	-
Negative Binomial	1.77 (1.13, 2.78)	527.91	1.73 (1.08, 2.77)	528.66
Zero-Inflated Poisson	1.71 (1.16, 2.52)	711.59	1.70 (1.17, 2.47)	699.10
Zero-Inflated Negative Binomial	1.87 (1.20, 2.92)	530.12	1.84 (1.16, 2.90)	530.91
Raw estimation	1.44 (1.01, 2.07)			

Table 8. Summary of the adjusted by baseline MQAI intervention effect by different count models for ET study

IQ study	Non-adjusted by baseline MQAI		Adjusted by baseline MQAI	
	Mean ratio (95% CI)	AIC	Mean ratio (95% CI)	AIC
Poisson	1.18 (0.49, 2.83)	906.59	1.18 (0.49, 2.84)	908.23
Quasi-Poisson	1.18 (0.49, 2.84)	-	1.18 (0.49, 2.83)	-
Negative Binomial	1.14 (0.48, 2.70)	350.59	1.16 (0.47, 2.88)	352.38
Zero-Inflated Poisson	1.37 (0.59, 3.20)	826.49	1.37 (0.58, 3.23)	830.04
Zero-Inflated Negative Binomial	1.23 (0.50, 2.98)	353.53	1.25 (0.49, 3.15)	357.24
Raw estimation	1.16 (0.48, 280)			

Table 9. Summary of the adjusted by baseline MQAI intervention effect by different count models for IQ study

As a secondary result, Figure 18 shows the effect of the intervention by study design. For both cross-sectional and interventional studies, the intervention increases the mean-per-year ratio by 1.97 (95% CI: 1.05 to 3.70) and 1.88 (95% CI: 1.19 to 2.98), respectively. The analysis for longitudinal and other types of studies point to a null effect: 1.03 (95% CI: 0.61 to 1.75) and 0.81 (95% CI: 0.28 to 2.38).

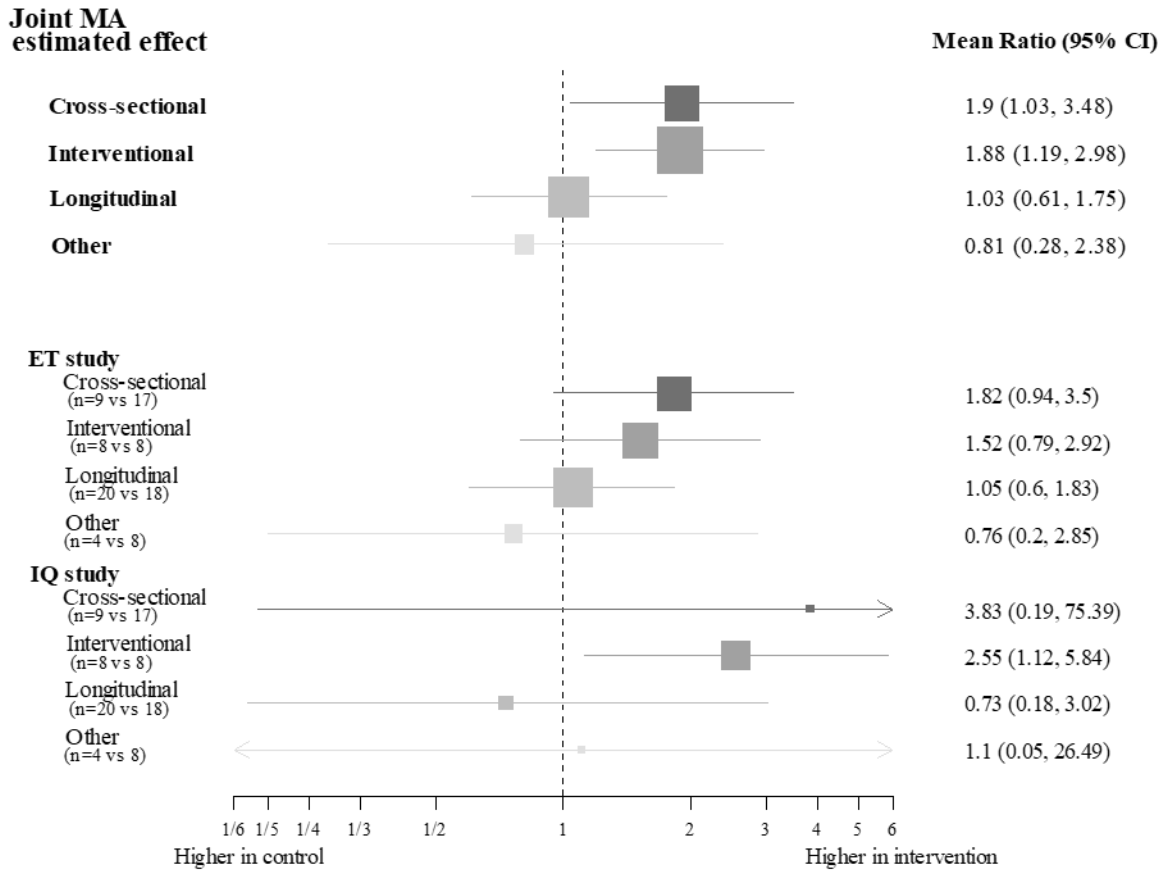


Figure 18. Citations-per-year mean ratio of SC additional intervention by study design

Figure 19 shows the MCF for the reference and the intervention groups. During the whole follow-up period, MCF is greater for the SC than for the R group. For example, the average NC experienced in 5 years in the SC group is 4.10 (95% CI: 2.97 to 5.23) versus 3.05 (95% CI: 2.40 to 3.70) citations per article. This figure reveals that the cumulative NC growth seems constant in the SC group, while this growth decrease with time for the R group, mainly after four years after publication.

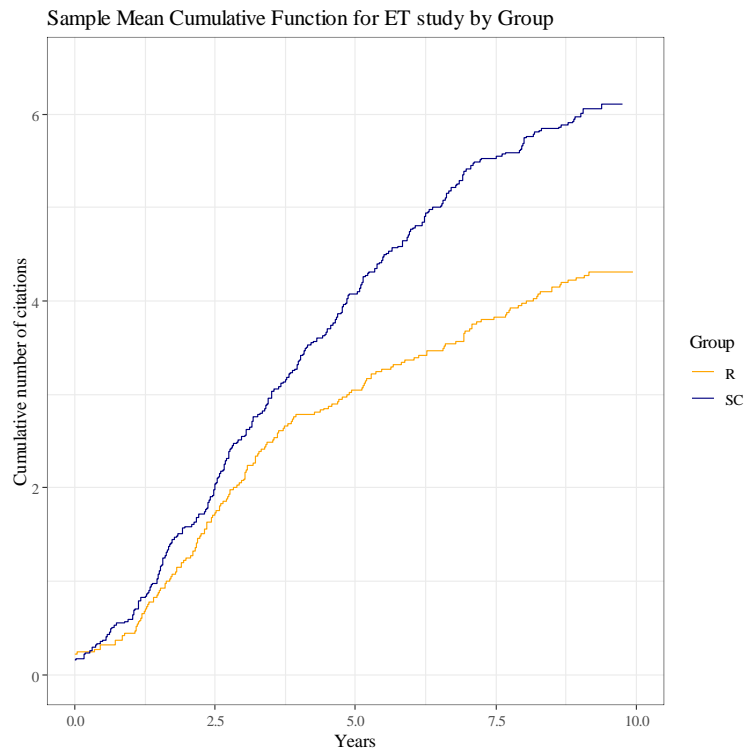


Figure 19. Mean cumulative function for the ET study by group

Figure 20 presents the MCF differences between the intervention and reference groups, with their random uncertainty measured as 95% CI. It shows that the intervention would increase in one citation per article within five years on average. This figure also reflects that the MCF difference grows over the years until the seventh year after publication, when the difference is stabilized. This difference is always in favor of the intervention group. As the MCF point estimate of the difference remains above zero, it indicates that the effect of the intervention persists over years.

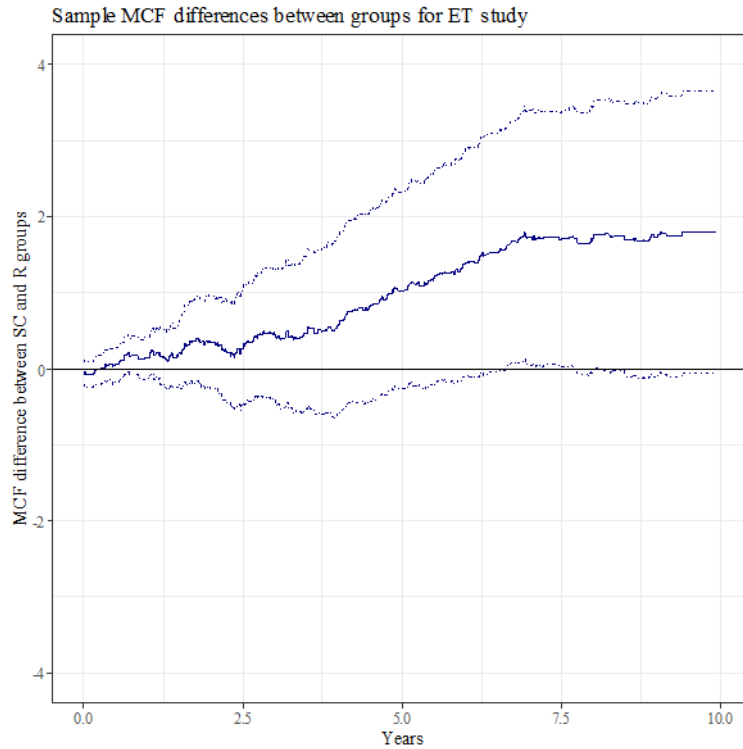


Figure 20. Mean cumulative function differences between groups for the ET study

We analyze the NC as recurrent events with a semi-parametric frailty model. We compute the frailty model with the different packages available in R software since it seems that there are no differences with the values of mean square error and bias in the simulation study (see Annex G). We exclude the `parfm` package since it does not converge neither in the simulation study nor in the data of study.

Baseline hazard ratio is non-parametric, and we assume that the frailty term follows a gamma distribution. Results shown below (Table 10, Figure 21 and Figure 22) are not adjusted for the baseline MQAI.

For the ET study, the hazard of receiving a citation for an article in the SC group at any time t compared to an R group article is 1.42 (95% CI: 1.00 to 2.03) according to the survival package. The estimation for the `frailtyEM` package is similar with a HR equals to 1.41 (95% CI: 1.00 to 2.02). The estimation for the `frailtypack` package is 1.39 (95% CI: 1.05 – 1.85). For all packages, the estimations suggest that the articles in the SC group have higher hazard to receive a citation.

	ET study		IQ study	
	HR (95% CI)	Variance of frailty (95% CI)	HR (95% CI)	Variance of frailty (95% CI)
<code>survival::coxph</code>	1.42 (1.00, 2.03)	0.52 (-)	1.16 (0.36, 3.69)	0.82 (-)
<code>frailtyEM::emfrail</code>	1.41 (1.00, 2.02)	0.51 (0.34, 0.77)	1.17 (0.62, 2.18)	0.84 (0.55, 1.25)
<code>frailtypack::frailtyPenal</code>	1.39 (1.05, 1.85)	0.28 (0.12, 0.44)	1.29 (0.79, 2.11)	0.71 (0.37, 1.05)

Table 10. Estimation of the hazard ratio comparing the expected number of citations between R and SC groups for ET and IQ studies.

Figure 21 shows the cumulative hazard function of R and SC groups for ET and IQ studies. Interpretation of this cumulative hazard is the expected number of citations at a certain time. For ET study, articles in SC group are expected to have higher NC at each time. For IQ study, the SC group also have higher expected NC although this difference is less strong and the confidence intervals are wider.

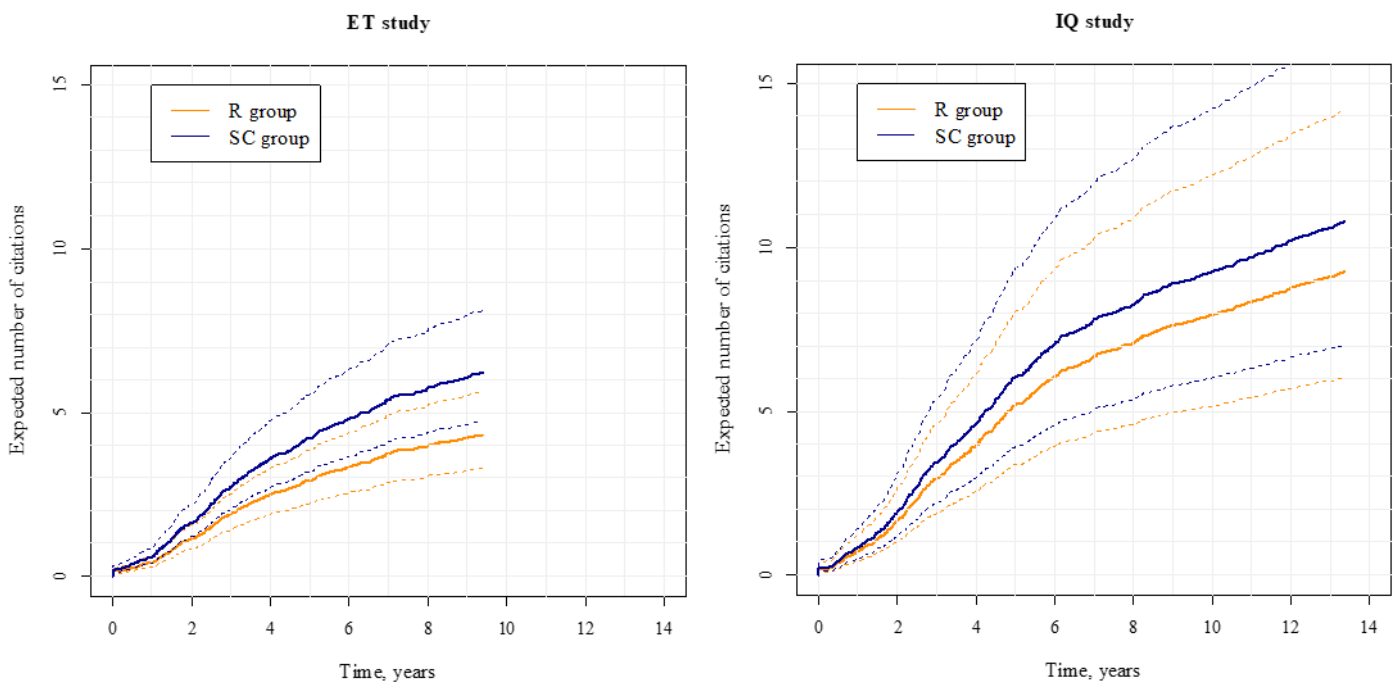


Figure 21. Comparison of the expected number of citations at a certain time between reference and intervention groups for ET and IQ studies

Variance of frailty for the frailtyEM package (see Table 10), which is the one that better MSE and bias values obtained in the simulation study (see Annex G), is 0.51 (95% CI: 0.34 to 0.77) for the ET study. Therefore, the hazard given that a citation occurs at time t is 51% higher than the hazard given that a citation does not occur at time t .

Figure 22 shows the estimated frailty distribution relying on a gamma density for the ET and IQ studies.

For the ET study, the quartiles are 0.48, 0.84 and 1.35. Without splitting by intervention group, articles with frailty at the first quartile (Q1) have 43% ($Q1/median - 1$) lower risk of obtaining a citation, and articles with frailty at the third quartile (Q3) have 61% ($Q3/median - 1$) higher risk of obtaining a citation than articles with median frailty. Clearly, the relationship between citations within each article have a substantial effect on the probability of obtaining a citation.

For both the R and SC groups, articles with frailty at Q1 have 34% lower risk of obtaining a citation than articles with median frailty. Articles with frailty at Q3 have 17% and 84% higher risk than articles with median frailty for R and SC groups respectively. Therefore, in SC group there are more articles with higher risk of citation.

For the IQ study, the quartiles are 0.34, 0.74 and 1.38. Without splitting by intervention group, articles with frailty at Q1 have 54% lower risk of obtaining a citation, and articles with frailty at Q3 have 87% higher risk of obtaining a citation than articles with median frailty. Again, the relationship between citations within each article have a substantial effect on the probability of obtaining a citation.

For the R and SC groups, articles with frailty at Q1 have 50% and 53% lower risk of obtaining a citation than articles with median frailty. Articles with frailty at Q3 have 86% and 40% higher risk than articles with median frailty for R and SC groups respectively.

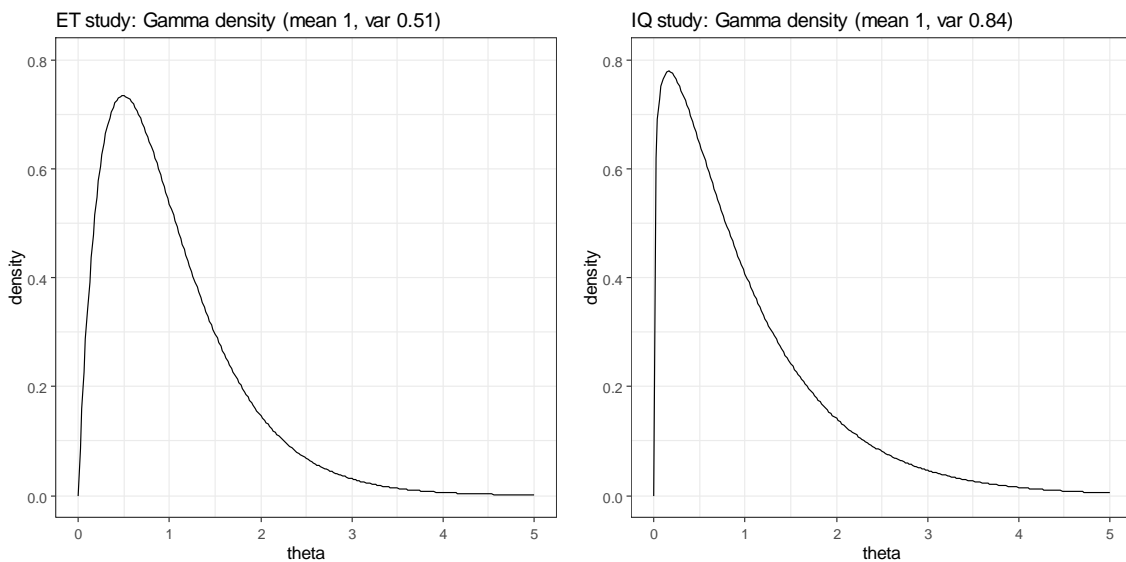


Figure 22. Gamma density of the estimated frailty distribution

For greater transparency, we present the main results updating the NC in July 2020. These results are in Annex I.

3.3. SEPARATE EFFECTS OF STATISTICAL REVIEWER AND REPORTING GUIDELINES

The separated effects are explored through the comparison of R+C versus S+SC (Statistician) and R+S versus C+SC (RG).

Figure 23 shows a negligible effect of the statistician, with a mean ratio equal to 1.03 (95% CI: 0.53 to 2.01); and a moderate effect of the RG, with mean ratio equal to 1.24 (95% CI: 0.64 to 2.40). Therefore, the intermediate IQ intervention groups showed concordant intermediate results, with more citations in the RG group.

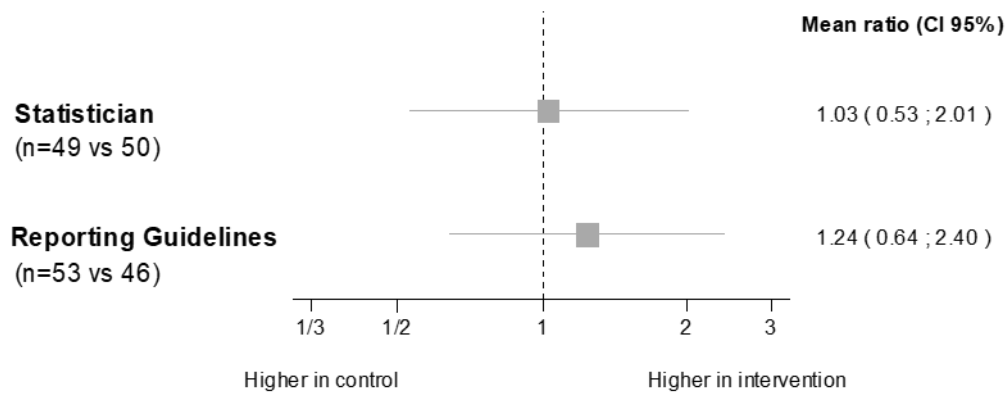


Figure 23. Citations-per-year mean ratio for Statistician and Reporting Guidelines effects

Chapter 4

DISCUSSION

The objective of this thesis is to evaluate the long-term impact of including a methodological reviewer (for example, a senior statistician) who is dedicated to looking for missing RG items in the peer-review panel of a journal. In the absence of previous exploratory studies designed for this purpose, we have conducted a proof of concept study. The results of this thesis will help to define the necessary parameters for carrying out future confirmatory studies to provide evidence of the benefits of such interventions.

4.1. MAIN FINDINGS

The intervention increases the number of citations by 40% (95% CI: 1 to 94%) when considering raw estimation without model assumptions. This effect increases to 61% (95% CI: 8 to 139%) with the negative binomial model, which is the parametric model that best fits NC. Considering citations as recurrent events and using a semi-parametric shared gamma frailty model, we find that the hazard of receiving a citation for an article in the SC group at any time t compared to an R group article is 1.41 (95% CI: 1.00 to 2.02).

If these findings are sustained, they justify the journal's cost and time expenditure (5) (4) (7) (8) (9). Interestingly, the number of papers with zero citations was higher in the intervention groups of both studies, which raises the possibility that greater transparency deters citations for some kinds of papers.

To the best of our knowledge, this is the first study that shows the potential association between completeness of reporting and higher citation counts due to the intervention of adding a methodological expert and reporting guidelines during the peer review process.

4.2. LIMITATIONS

Although the number of citations is considered by some authors to be an indicator of a paper's influence (143) (144) (145) (146) (147), some have argued that “citation counts are an indicator more of impact than of quality or importance” (148); thus, we should not conflate the number of citations with research quality (149) (150). We consider the NC impartial and fair, as it is naturally free from the risk of evaluation bias.

There are factors that can influence citations counts such as the online social media attention, some investigations have showed correlation between NC and the article level metrics (151) (152). Our meta-analysis combined the intervention effect of two studies with articles from two periods with different degrees of social media appearance. We ignore the effect of social media or other factors on NC, further studies are needed to investigate this issue.

Behind the IQ study lies considerable uncertainty, including the risk of bias due to attrition. The ET study has higher weight when estimating the combined effects; therefore, our interpretation mainly follows the ET results in that: when a methodologist formally uses RGs at the end of the editorial phase following peer review, this leads to an increase in the papers' potential scientific value. This interpretation assumes that all added citations are “positive” in the sense that they contribute to the body of knowledge.

As the ET and IQ studies were originally designed to test those hypotheses in a different outcome, our results should be interpreted as suggestive and need to be replicated.

Our next important limitation pertains to the fact that the results rely on just one journal that is not top-quality (Medicina Clínica, Barcelona), and they therefore cannot be transported to top-tier journals where those interventions have probably already been implemented. According to the Scimago Journal & Country Rank website, journals with Impact Factor ≥ 10 account for just 1% (15259 out of 1528749 articles published in 2016) of biomedical scientific production. Thus, our focus is not on the top-quality journals but on second-tier journals that could benefit from the intervention.

It is essential to interpret these results according to the exploratory nature of this extended follow-up study. First, we did not have enough advance information to know the fit between our data and the statistical models. Second, and more importantly, we had neither previous studies to sustain the

hypothesis nor a sample size rationale to guarantee any desired power for testing this hypothesis. Therefore, in keeping with the American Statistical Association (ASA) statement on p-value (153), we should not interpret the results of any hypothesis test. Accordingly, we should also not be concerned about whether or not the 95% confidence intervals (CI) include the neutral value of 1, because there is no such previous hypotheses. However, as we stated prior to the data collection that our objective is to estimate the effects of S and/or C interventions on the number of citations, selective outcome reporting is of no concern.

In order to hinder the reproducibility crisis that many authors claim (154) in biomedical research, we provide a Shiny application with R Software that shows the data of the study and the results obtained, all of which are available to anyone browsing the internet. With this initiative, we attempt to demonstrate that methods currently exist for promoting reproducibility so that other researchers can reproduce the analysis of a study and replicate the effect that was originally reported. It is important that scientific research focus on preventing falsifiability (154).

Apart from this manuscript, the process of carrying out this thesis led to our publishing an article (see Annex J) presenting the positive results of this post-hoc study. It reported the same methodology for count data using the NC that was collected up to 31 December 2016 as the main variable. This article was well-received by the research community and obtained six citations in the first year after publication: one citation found in WoS and five in CrossRef. [Online attention](#) measured by Altmetric calculated a score of 42 points within one year after publication, placing it in the 94th percentile (ranked 15,309th of the 264,209 tracked articles of a similar age in all journals). This attention is basically due to the 80 tweeters and 25 Mendeley mentions in the online community.

4.3. FUTURE RESEARCH

The implementation of this proof of concept study has opened up various lines of future research.

With the obtained results, we can design a confirmatory long-term impact study to verify the positive intervention effect. This future study will require two phases: the first will implement the intervention on articles in a second-tier journal; and the second phase will take place after a few years, when the effect of the intervention will be evaluated by analyzing the NC obtained for these articles. The results in Figure 20 make us consider a follow-up of around 7 years to estimate the long-term effect, since this figure shows that, after the 7th year after publication, the differences between the cumulative numbers of citations stabilize.

The secondary objectives for the confirmatory study could be to examine the consistency of the results using different sources of NC (WoS, Google Scholar, and Scopus, among others) and also examining the impact in social media by means of altmetrics.

The ET and IQ studies showed a positive relationship of the intervention, as measured by MQAI. While in the exploratory analysis of 2016 we found a relationship between MQAI and NC, this thesis shows the positive effect of the intervention on NC. These relationships make it challenging to determine the real cause of the long-term impact as measured by NC. Therefore, we propose to approach this challenge by using mediation analysis.

4.4. CONCLUSIONS

We have definitively shown that randomized studies are feasible for improving scientific impact, as measured by the number of WoS citations.

As the implementation costs of RG are low relative to the benefits, we strongly urge journals to perform their own studies to ascertain whether scientific impact increases by adhering to reporting guidelines and involving statisticians or methodological experts in the editorial process.

Our findings indicate that the citation counts of research papers increased by 40% (95% CI: 1% to 94%) after introducing into the editorial process a senior methodologist that uses the publication guidelines. As our studies were originally designed to test those hypotheses on a different outcome, our results should be interpreted as exploratory and need to be replicated. We invite journals to perform their own studies to ascertain whether scientific repercussion increases by adhering to reporting guidelines and further involving statisticians in the editorial process.

Chapter 5

BIBLIOGRAPHY

1. Gutiérrez J, Carrasco M. Gestión de la investigación biomédica. Reflexiones sobre la ciencia en España. El caso particular de la biomedicina. Madrid : Fundación Lilly, 2003. Vol. 137. 137-66.
2. Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica, 2000-2003. Presidencia del Gobierno: Oficina de Ciencia y Tecnología, 1999.
3. Nass SJ, Levit LA, Gostin LO. (2009). Beyond the HIPPA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington (DC): National Academies Press.
4. Chalmers I, Glasziou P. (2009). Avoidable waste in the production and reporting of research evidence. *Lancet*. 374 (9683); 86-89. doi:10.1016/S0140-6736(09)60329-9.
5. Macleod M, Michie S, Roberts I, et al. (2014). Biomedical research: increasing value, reducing waste. *Lancet*. 383 (9912); 101-104. doi:10.1016/S0140-6736(13)62329-6.
6. Chalmers I, Bracken MB, Djulbegovi B, et al. (2014). How to increase value and reduce waste when research priorities are set. *Lancet*. 383 (9912); 156-165.
7. Ioannidis JPA, Greenland S, Hlatky MA, et al. (2014). Increasing value and reducing waste in research design, conduct and analysis. *Lancet*. 383 (9912); 166-175. doi:10.1016/S0140-6736(13)62227-8.
8. Al-Shahi Salman R, Beller E, Kagan J, et al. (2014). Increasing value and reducing waste in biomedical research regulation and management. *Lancet*. 383 (9912); 176-185. doi:10.1016/S0140-6736(13)62297-7.
9. Chan A-W, Song F, Vickers A, et al. (2014). Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 383 (9913); 257-266. doi:10.1016/S0140-6736(13)62296-5.

10. Glasziou P, Altman DG, Bossuyt P, et al. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 383 (9913); 267-276.
11. Moher D, Glasziou P, Chalmers I, et al. (2016). Increasing value and reducing waste in biomedical research: who's listening. *Lancet*. 387 (10027); 1573-1586. doi:10.1016/S0140-6736(15)00307-4.
12. Peer review process and editorial decision making at journals . Part 02 - Status descriptions in journal submission systems . Editage Insights, 2013.
13. Rothwell PM, Martyn CN. (2000). Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain*. 123 (9); 1964-1969. doi:10.1093/brain/123.9.1964.
14. Goodman SN, Berlin J, Fletcher SW, Fletcher RH. (1994). Manuscript quality before and after peer review and editing at *Annals of Internal Medicine*. *Internal Annals of Medicine*. 121 (1); 11-21. doi:10.7326/0003-4819-121-1-199407010-00003.
15. Bruce R, Chauvin A, Trinquart L, Ravaud P, Boutron I. (2016). Impact of Interventions to Improve the Quality of Peer Review of Biomedical Journals: A Systematic Review and Meta-analysis. *BMC Medicine*. 14 (1): 85. doi:10.1186/s12916-016-0631-5.
16. Jefferson T, Rudin M, Brodney Folse S, Davidoff F. (2007). Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database Systematic Review*. 2: MR000016. doi:10.1002/14651858.MR000016.pub3.
17. Shamseer L, Stevens A, Skidmore B, et al. (2012). Does journal endorsement of reporting guidelines influence the completeness of reporting of health research? A systematic review protocol. *Systematic Reviews*. 1: 24. doi:10.1186/2046-4053-1-24.
18. Moher D, Hopewell S, Schulz KF, et al. (2012). CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 10 (1): 28-55. doi:10.1016/j.ijsu.2011.10.001.
19. Begg C, Cho M, Eastwood S, et al. (1996). Improving the quality of reporting of randomized controlled trials, the CONSORT statement. *JAMA*. 276 (8): 637-639. doi:10.1001/jama.276.8.637.
20. Marusic A, Gasparyan A, Kitas GD. (2013). Promoting transparent and accurate reporting of research studies in rheumatology: Endorsement of reporting guidelines in rheumatology journals. *Seminars in Arthritis and Rheumatism*. 43 (2): 288-291. doi:10.1016/j.semarthrit.2013.01.005.
21. Schoonbaert D, Roelants G. (1996). Citation analysis for measuring the value of scientific publications: quality assessment tool or comedy of errors? *Tropical medicine and International health*. 1 (6): 739-752. doi:10.1111/j.1365-3156.1996.tb00106.x.

22. Yadav CP, Sreenivas V, Khan MA, Pandey RM. (2018). An overview of statistical models for recurrent events analysis: a review. *Epidemiology*. 8: 354. doi: 10.4172/2327-4972.1000354.
23. Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication. (2010). *J Pharmacol Pharmacother*. 1 (1): 42-58.
24. Smith R. (2006). Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*. 99 (4): 178–182. doi:10.1258/jrsm.99.4.178.
25. Jefferson T, Alderson P, Wager E, Davidoff F. (2002). Effects of editorial peer review. *JAMA*. 287(21):2784-6. doi:10.1001/jama.287.21.2784.
26. Kelly J, Sadeghieh T, Adeli K. (2014). Peer review in scientific publications: benefits, critiques, & a survival guide. *EJIFCC*. 25 (3): 227-243.
27. Superchi C, González JA, Solà I, Cobo E, Hren D, Boutron I. (2019). Tools used to assess the quality of peer review reports: a methodological systematic review. *BMC Med Res Methodol*. 19: 48. doi: 10.1186/s12874-019-0688-x.
28. Sense about Science. [Online] <https://senseaboutscience.org/activities/peer-review-survey-2019/>.
29. Schroter S, Black N, Evans S, Carpenter J, Godlee F, Smith R. (2004). Effects of training on quality of peer review: randomised controlled trial. *BMJ*. 328 (7441): 673. doi:10.1136/bmj.38023.700775.AE.
30. Patel J. (2014). Why training and specialization is needed for peer review: a case study of peer review for randomized controlled trials. *BMC Med*. 12:128. doi: 10.1186/s12916-014-0128-z.
31. Lovejoy T, Revenson T, France C. (2011). Reviewing manuscripts for peer-review journals: a primer for novice and seasoned reviewers. *Annals of behavioural medicine*. 42 (1): 41275. doi:10.1007/s12160-011-9269-x.
32. Baxt WG, Waeckerle JF, Berlin JA, et al. (1998). Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. *Annals of Emergency Medicine*. 32 (3 pt 1): 310–317. doi:10.1016/s0196-0644(98)70006-x.
33. Shashok K. (2008). Content and communication: How can peer review provide helpful feedback about the writing. *BMC Medical Research Methodology*. 8: 3. doi: 10.1186/1471-2288-8-3.
34. Kravitz RL, Franks P, Feldman MD, et al. (2010). Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS One*. 5 (4): e10072. doi:10.1371/journal.pone.0010072.
35. Henderson M. (2010). Problems with peer review. *BMJ*. 340: c1409. doi:10.1136/bmj.c1409.
36. Yaffe MB. (2009). Re-reviewing peer review. *Scielo Signal*. 2 (85): eg11. doi:10.1126/scisignal.285eg11.

37. Cobo E, González JA. (2016). Taking advantage of unexpected WebCONSORT results. *BMC Medicine*. 14 (1): 204. doi:10.1186/s12916-016-0758-4.
38. Levitt JM, Thelwall M. (2011). A combined bibliometric indicator to predict article impact. *Information processing & management*. 47 (2): 300-308. doi: 10.1016/j.ipm.2010.09.005.
39. Hopewell S, Boutron I, Altman DG, et al. (2016). Impact of a web-based tool (WebCONSORT) to improve the reporting of randomised trials: results of a randomised controlled trial. *BMC Med*. 14 (1): 199. doi:10.1186/s12916-016-0736-x.
40. Stahel PF, Moore EE. (2014). Peer review for biomedical publication: we can improve the system. *BMC Medicine*. 12 (1): 179. doi:10.1186/s12916-014-0179-1.
41. Editors, Council of Science. Policy on Responsibilities and Rights of Editors of Peer-Reviewed Journals. http://www.councilscienceeditors.org/services/draft_approved.cfm#Authors. [Online] [09/02/2016]
42. Altman DG. (1982). Statistics in medical journals. *Statistical Medicine*. 1 (1): 59-71. doi:10.1002/sim.4780010109.
43. Altman DG. (1994). The scandal of poor medical research. *BMJ*. 308 (6924): 283-284. doi:10.1136/bmj.308.6924.283.
44. Herbison P. (2007). How to make your article more acceptable for the statistical reviewer (Editorial). *Neurolog Urology*. 26 (3): 318-22. doi:10.1002/nau.20346.
45. Savik K. (2006). Reporting p-values. *J Wound Ostomy Continence Nurs*. 33 (3): 247-251. doi:10.1097/00152192-200605000-00004.
46. Bailar JC 3rd, Mosteller F. (1988). Guidelines for statistical reporting in article for medical journals. Amplifications and explanations. *Ann Intern Med*. 108 (2): 266-73. doi:10.7326/0003-4819-108-2-266.
47. Haller KB, Holditch-Davies D. (2000). Guidelines for statistical reporting. *JOGNN*. 29 (2): 121. doi:10.1111/j.1552-6909.2000.tb02030.x.
48. Murray GD. (1988). The task of a statistical referee. *Br J Surg*. 75 (7): 664-667. doi:10.1002/bjs.1800750714.
49. Sosa JA, Mehta P, Thomas DC, et al. (2009). Evaluating the surgery literature: can standardization today predict manuscript impact tomorrow? *Ann Surg*. 250 (1): 152-158. doi:10.1097/SLA.0b013e3181ad8905.
50. Wang X, Chen Y, Yang N, et al. (2015). Methodology and reporting quality of reporting guidelines: systematic review. *BMC Medical Research Methodology*. 15: 74. doi:10.1186/s12874-015-0069-z.

51. Simera I, Moher D, Hirst A, et al. (2010). Transparent and accurate reporting increases reliability, utility and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Medicine*. 8: 24. doi: 10.1186/1741-7015-8-24.
52. McIntosh LD, Juehne A, Vitale CRH, et al. (2017). Repeat: a framework to assess empirical reproducibility in biomedical research. *BMC Medical Research Methodology*. 17 (1): 143. doi:10.1186/s12874-017-0377-6.
53. Park IU, Peacey MW, Munafò MR. (2014). Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*. 506 (7486): 93-96. doi:10.1038/nature12786.
54. Triaridis S, Kyrgidis A. (2010). A peer review and journal impact factor: the two pillars of contemporary Medical publishing. *Hippokratia*. 14 (Suppl 1): 5-12.
55. Shamseer L, Stevens A, Skidmore B, et al. (2012). Does journal endorsement of reporting guidelines influence the completeness of reporting of Health research? A systematic review protocol. *Systematic Reviews*. 1: 24. doi:10.1186/2046-4053-1-24.
56. Simera I, Altman DG, Moher D, et al. (2008). Guidelines for reporting health research: the EQUATOR network's survey of guidelines authors. *PLoS Med*. 5 (6): e139. doi:10.1371/journal.pmed.0050139.
57. Enhancing the QUALity and Transparency Of health Research. equator network. [Online] [06/11/ 2017]. <https://www.equator-network.org/>.
58. Sharp M, Tokalic R, Gómez C, Wager E, Altman DG, Hren D. (2019). A cross-sectional bibliometric study showed suboptimal endorsement rates of STROBE and its extensions. *Journal of clinical epidemiology*. 107: 42-50. doi:10.1016/j.jclinepi.2018.11.006.
59. Sharp M, Bertixxolo L, Rius R, Wager E, Gómez G, Hren D. (2019). Using the STROBE statement: survey findings emphasized the role of journals in enforcing reporting guidelines. *Journal of Clinical Epidemiology*. 116: 26-35. doi:10.1016/j.jclinepi.2019.07.019.
60. Moher D, Plint AC, Altman DG, et al. (2010). Consolidated Standards of Reporting Trials (CONSORT) and the Quality of Reporting of Randomized Controlled Trials (Protocol). *Cochrane Database of Systematic Reviews*: MR000030. doi: 10.1002/14651858.MR000030.
61. T von Elm E, Altman DG, Egger M, et al. (2008). The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *J Clin Epidemiol*. 61 (4): 344-349. doi:10.1016/j.jclinepi.2007.11.008.
62. Moher D, Liberati A, Tetzlaff J, Altman DG and PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 6 (7): e1000097. doi:10.1371/journal.pmed.1000097.

63. Chan AW, Tetzlaff JM, Altman DG, et al. (2013). SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Annals of Internal medicine*. 158 (3): 200-207.
64. Moher D, Shamseer L, Clarke M, et al. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Syst Rev*. 4 (1): 1. doi:10.1186/2046-4053-4-1.
65. Bossuyt PM, Reitsma JB, Bruns DE, et al. (2003). Standards for reporting of diagnostic accuracy. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 326 (7379): 41-44. doi:10.1136/bmj.326.7379.41.
66. Collins GS, Reitsma JB, Altman DG, Moons KG. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis: the TRIPOD statement. *BMJ*. 350: g7594. doi:10.1136/bmj.g7594.
67. Gagnier JJ, Kienle G, Altman DG, Moher K, Sox H, Riley D. (2013). The CARE guidelines: consensus-based clinical case reporting guideline development. *Glob Adv Health Med*. 2 (5)-. 38-43. doi:10.7453/gahmj.2013.008.
68. Brouwers MC, Kerkvliet K, Spithoff K. (2016). The AGREE reporting checklist: a tool to improve reporting of clinical practice guidelines. *BMJ*. 352: i1152. doi:10.1136/bmj.i1152.
69. Chen Y, Yang K, Marušić A, et al. (2017). A reporting tool for practice guidelines in health care: the RIGHT statement. *Ann Intern Med*. 166 (2): 128-132. doi:10.7326/M16-1565.
70. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. (2014). Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. 89 (9): 1245-1251. doi:10.1097/ACM.0000000000000388.
71. Tong A, Sainsbury P, Craig J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *nt J Qual Health Care*. 19 (6): 349-357. doi:10.1093/intqhc/mzm042.
72. Kilkeny C, Browne W, Cuthill IC, Emerson M, Altman DG y Group, NC3Rs Reporting Guidelines Working. (2010). Animal research: reporting in vivo experiments - the ARRIVE guidelines. *Br J Pharmacol*. 160 (7): 1577-1579. doi:10.1111/j.1476-5381.2010.00872.x.
73. Ogrinc G, Davies L, Goodman D, Batalden P, Davidoff F, Stevens D. (2016). SQUIRE 2.0 (Standards for Quality Improvement Reporting Excellence): revised publication guidelines from a detailed consensus process. *BMJ Qual Saf*. 25 (2): 986-992. doi:10.1136/bmjqs-2015-004411.
74. Husereau D, Drummond M, Petrou S, et al. (2013). Consolidated Health economic evaluation reporting standards (CHEERS) statement. *Value Health*. 16 (2): e1-e5. doi:10.1016/j.jval.2013.02.010.

75. D Moher, K Schulz, DG Altman. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Lancet*. 357 (9263): 1191-1194.
76. Altman DG, Schulz KF, Moher D, et al. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 134 (8): 663-694. doi:10.7326/0003-4819-134-8-200104170-00012.
77. Schulz KF, Altman DG, Moher D. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 340: c332. doi:10.1136/bmj.c332.
78. Plint AC, Moher D, Morrison A, et al. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Med J Aust*. 185 (5): 263-267.
79. Turner L, Moher D, Shamseer L, et al. (2011). The influence of CONSORT on the quality of reporting of randomized controlled trials: an updated review. *Trials*. 12 (Suppl 1): A47. doi:10.1186/1745-6215-12-S1-A47.
80. Richards D. (2007). Little evidence to support the use of editorial peer review to ensure quality of published research. *Evid Based Dent*. 8 (3): 88-89. doi:10.1038/sj.ebd.6400516.
81. Blanco D, Altman D, Moher D, Boutron I, Kirkham JJ, Cobo E. (2019). Scoping review on interventions to improve adherence to reporting guidelines in health research. *BMJ Open*. 9 (5): e026589. doi:10.1136/bmjopen-2018-026589.
82. Repiso R, Castillo-Esparcia A, Torres-Salinas D. (2019). Altmetrics, alternative indicators for Web of Science Communication studies journals. *Scientometrics*. doi: 10.1007/s11192-019-03070-7.
83. Egge L, Rousseau R. (1990). Introduction to informetrics. Amsterdam : Quantitative methods in library, documentation and information science. Elsevier.
84. Buela-Casal G, Zych I. (2010). Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals. *Psicothema*. 22 (2): 270-276.
85. Callahan M, Wears RL, Weber E. (2002). Journal prestige, publication bias and other characteristics associated with citation of published studies in peer-reviewed journals. *Journal of the American Medical Association*. 287 (21): 2847-2850. doi:10.1001/jama.287.21.2847.
86. CJC Jabbour, ABLD Jabbour, JHC Oliveira. (2013). The perception of brazilian researchers concerning the factors that influence the citation of their articles: A study in the field of sustainability. *Serials Review*. 39 (2): 93-96.
87. Patterson MS, Harris S. (2009). The relationship between reviewers' quality-scores and number of citations for papers published in the journal physics in medicine and biology from 2003-2005. *Scientometrics*. 80 (2): 343-349.

88. Stremersch S, Verniers I, Verhoef PC. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing*. 71 (3): 171-193. doi: 10.1509/jmkg.71.3.171.
89. Bornmann L, Bowman BF, Bauer J, et al. (2012). Standards for applying bibliometrics to the evaluation of research institutes in the Natural Sciences. *Zeitschrift für Evaluation*. 11 (2): 233-260.
90. Garfield E. (2006). The history and meaning of the journal impact factor. *Journal of the American Medical Association*. 295 (1): 90-93. doi:10.1001/jama.295.1.90.
91. Tahamtan I, Safipour Afshar A, Ahmadvadeh K. (2016). Factors affecting number of citations: a comprehensive review of the literature. *Scientometrics*. 107 (2016): 1195-1225. doi: 10.1007/s11192-016-1889-2.
92. Chakraborty T, Kumar S, Goyal P, Ganguly N, Mukherjee M. (2014). Towards a stratified learning approach to predict future citation counts. *Proceedings of the ACM/IEEE joint conference on digital libraries*. doi: 10.1109/JCDL.2014.6970190. .
93. Hilmer C, Lusk J. (2009). Determinants of citations to the agricultural and applied economics association journals. *Review of Agricultural Economics*. 31 (4): 677-694. doi: 10.2307/40588523.
94. Garner J, Porter AL, Newman NC. (2014). Distance and velocity measures: Using citations to determine breadth and speed of research impact. *Scientometrics*. 100 (3): 687-703.
95. Harwood N. (2008). Citers' use of citees' names: Findings from a qualitative interview-based study. *J Am Soc Inf Sci*. 59: 1007-1011. doi:10.1002/asi.20789.
96. Aksnes DW, Sivertsen G. (2004). The effect of highly cited papers on national citation indicators. *Scientometrics*. 59 (2): 213-224. doi: 10.1023/B:SCIE.0000018529.58334.eb.
97. Bernstein J, Gray C. (2012). Content factor: a measure of a journal's contribution to knowledge. *Plos One*. 7 (7): e41554. doi:10.1371/journal.pone.0041554.
98. Scimago Journal & Country Rank. [Online][09/05/2018] <https://www.scimagojr.com/journalrank.php>.
99. Markusova VA, Mindeli LE, Bogorov VG, Libkind AN. (2018). An altmetric as an indicator of a publication's scientific impact. *Herald of the Russian Academy of Science*. 88 (5): 394-400. doi: 10.1134/S1019331618040123.
100. Scopus IL, Sedighi M. (2019). The Role of Social media in assessing the impact of research. *Iranian Journal of Information processing and management*. 34 (2): 765-792.
101. Mullins CH, Boyd CJ, Corey BL. (2020). Examining the Correlation Between Altmetric Score and Citations in the General Surgery Literature. *J Surg Res*. 248: 159-164. doi:10.1016/j.jss.2019.11.008.

102. Martín-Martín A, Orduna-Malea E, Thelwall M, Delgado López-Cózar E. (2019). Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. *Journal of Informetrics*. 12 (4): 1160-1177. doi: 10.1016/j.joi.2018.09.002.
103. Higgins JPT, Green S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* version 5.1.0 [updated March 2011]. Wiley-Blackwell.
104. Van Rooyen S, Black N, Godlee F. (1999). Development of the review quality instrument (RQI) for assessing peer reviews of manuscripts. *J Clin Epidemiol*. 52 (7): 625-629. doi: 10.1016/S0895-4356(99)00047-5.
105. Callahan M, Baxt W, Waeckerle J, et al. (1998). The reliability of editors' subjective quality ratings of manuscript peer reviews. *JAMA*. 280 (3): 229-231. doi:10.1001/jama.280.3.229.
106. Okike K, Hug KT, Kocher MS, Leopold SS. (2016). Single-blind vs Double-blind Peer Review in the Setting of Author Prestige. *JAMA*. 316 (12): 1351-1316. doi:10.1001/jama.2016.11014.
107. Cobo E, Selva-O'Callaghan A, Ribera JM, et al. (2007). Statistical reviewers improve reporting in Biomedical articles: a randomized trial. *Plos One*. 2 (3): e332. doi:10.1371/journal.pone.0000332.
108. Cobo E, Cortés J, Ribera JM, et al. (2011). Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomized trial. *BMJ*. 343: d6783. doi: 10.1136/bmj.d6783.
109. Arnau C, Cobo E, Ribera JM, Cardellach F, Selva A, Urrutia A. (2003). Effect of statistical review on manuscript quality in *Medicina Clínica* (Barcelona): a randomized study. *Medicina Clínica*. 121 (18): 690-694. doi: 10.1016/s0025-7753(03)74064-0.
110. Web of Science. [Online] <https://webofknowledge.com>.
111. Bosch F, Guardiola E and Fundación Esteve working group. Abridged checklist for the evaluation of basic biomedical research papers. *Medicina Clínica*. 121 (6): 228-230. doi:10.1016/s0025-7753(03)73913-x.
112. Moher D, Cook DJ, Eastwood S, et al. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Quality of Reporting of Meta-analyses*. *Lancet*. 354 (9193): 1896-1900. doi:10.1016/s0140-6736(99)04149-5.
113. Mora i Ripoll R. (1999). How to improve the statistical quality of articles presented to biomedical journals: a checklist for authors. *Med Clin (Barc)*. 113 (4): 138-49.
114. Zeileis A, Kleiber C, Jackman S. (2008). Regression models for count data in R. *Journal of statistical software*. 27 (8): 1-25. doi: 10.18637/jss.v027.i08.

115. Cameron AC, Trivedi PK. (1998). *Regression Analysis of Count Data*. Cambridge University. doi:10.1017/CBO9781139013567.
116. Long JS, Freese J. (2001). *Regression models for categorical dependent variables using Stata*. Stata corporation. ISBN 1-881228-62-2.
117. Cameron AC, Trivedi PK. (2005). *Microeconometrics: methods and applications*. Cambridge University Press. doi: 10.1017/CBO9780511811241.
118. Chambers JM, Hastie TJ. (1992). *Statistical models in S*. Chapman & Hall.
119. Gardner W, Mulvey EP, Shaw EC. (1995). Regression analyses of counts and rates: poisson, over dispersed Poisson, and negative binomial models. *Psychol Bull.* 118 (3): 392.
120. Ver Hoef JM, Boveng PL. (2007). Quasi-Poisson vs. negative binomial regression: how should we model over dispersed count data? *Ecology.* 88 (11): 2766-72.
121. Venables WN, Ripley BD. (2002). *Modern Applied Statistics with S*. New-York : Springer-Verlag, 4th edition.
122. J, Mullahy. (1986). Specification and testing of some modified count data models. *Journal of Econometrics.* 33 (3): 341-365. doi: 10.1016/0304-4076(86)90002-3.
123. D, Lambert. (1992). Zero-Inflated Poisson Regression, with an application to defects in manufacturing. *Technometrics.* 34: 1-14. doi: 10.1080/00401706.1992.10485228.
124. Burnham KP, Anderson DR. (2002). *Model selection and multimodel inference: a practical information theoretic approach*. Second edition. Springer-Verlag.
125. D, Lakens. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychology.* 4: 863. doi:10.3389/fpsyg.2013.00863.
126. Andersen PK, Gill RD. (1982). Cox's Regression model for counting processes: a large sample study. *Annals of statistics.* 10 (4): 1100-1120. doi: 10.1214/aos/1176345976.
127. Prentice RL, Williams BJ, Peterson AL. (1981). On the regression analysis of multivariate failure time data. *Biometrika.* 68 (2): 373-379. doi: 10.1093/biomet/68.2.373.
128. Wei LJ, Lin DY, Weissfeld L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *JASA.* 84 (408): 1065-1073. doi: 10.1080/01621459.1989.10478873.
129. O, Siddiqui. (2009). Statistical methods to analyse adverse events data of randomized clinical trials. *J Biopharm Stat.* 19 (5): 889-899. doi:10.1080/10543400903105463.

130. W, Nelson. (2003). Recurrent events data analysis for product repairs, disease recurrences, and other applications. *Technometrics*. 45 (3): 262-264. doi: 10.1198/004017003000000087.
131. Viljanen M, Airola A, Majanoja AM, et al. (2017). Measuring Player Retention and Monetization using the Mean Cumulative Function. *IEEE Transactions on Games*. doi: 10.1109/TG.2020.2964120.
132. Cook RJ, Lawless JF. (2007). The statistical analysis of recurrent events. *Statistics for biology and health*. doi: 10.1007/978-0-387-69810-6.
133. Xu Y, Cheung YB, Lamd KF, Milligane P. (2012). Estimation of summary protective efficacy using a frailty mixture model for recurrent event time data. *Stat in Med*. 31 (29): 4023-4039. doi: 10.1002/sim.5458.
134. Duchateau L, Janssen P. (2008). *The Frailty Model*. New York : Springer-Verlag. doi: 10.1007/978-0-387-72835-3.
135. Monaco JV, Gorfine M, Hsu L. (2018). General semiparametric shared frailty model: estimation and simulation with frailtySurv. *Journal of statistical software*. 86: 4. doi:10.18637/jss.v086.i04.
136. Wienke A. (2010). *Frailty models in survival analysis*. New York : Chapman and Hall. doi: 10.1201/9781420073911.
137. Hougart P. (2000). *Analysis of multivariate survival data*. New York : Springer-Verlag. doi: 10.1007/978-1-4612-1304-8.
138. Clayton DG. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*. 65 (1): 141-151. doi:10.2307/2335289.
139. Munda M, Rotolo F, Legrand C. (2012). parfm: Parametric frailty models in R. s.l. : *Journal of Statistical software*. 51 (11): 1-20. doi: 10.18637/jss.v051.i11.
140. Therneau TM. (2018). survival: a package for survival analysis in S. R package version 2.42-3. <https://CRAN.R-project.org/package=survival>.
141. Therneau TM. (2018). coxme: Mixed effects cox models. R package version 2.2-7. <https://CRAN.R-project.org/package=coxme>.
142. Wallach JD, Boyack KW, Ioannidis JPA. (2018). Reproducible research practices, transparency, and open access data in the biomedical literature. *PLoS Biol*. 16 (11): 2015–2017.
143. Ferrante di Ruffano L, Dinnes J, Taylor-Philips S, et al. (2017). Research waste in diagnostic trials: a method review evaluating the reporting of test-treatment interventions. *BMC Medical Research Methodology*. 17: 32.

144. MacRoberts M, MacRoberts B. (1989). Problems of citation analysis: A critical review. *Journal of the American Society for Information Science*. 40: 342-349.
145. Subelj L, Fiala D, Bajec M. (2014). Network-based statistical comparison of citation topology of bibliographic databases. *Nature*. 4: 6496.
146. Yao L, Wei T, Zeng A, Fan Y, Di Z. (2014). Ranking scientific publications: the effect of nonlinearity. *Nature*. 4: 6663.
147. Zhang Z, Van Poucke S. (2014). Citations for randomized controlled trials in sepsis literature: the halo effect caused by journal impact factor. *Plos One*. 12 (1): e0169398.
148. Martin BR. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*. 36 (3): 343-362.
149. Nieminen P, Carpenter J, Rucker G, Schumacher M. (2006). The relationship between quality of research and citation frequency. *BMC Medical Research Methodology*. 6: 42.
150. Seglen P. (1992). The skewness of science. *Journal of the American Society for Information Science*. 43 (9): 628-638.
151. Luc JGY, Archer MA, Arora RC, Bender EM, Blitz A, et al. (2020). Social Media Improves Cardiothoracic Surgery Literature Dissemination: Results of a Randomized Trial. *Ann Thorac Surg*. 109(2): 589-595. doi: 10.1016/j.athoracsur.2019.06.062..
152. G., Eysenbach. (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res*. 13(4): e123. doi: 10.2196/jmir.2012.
153. Wassersetin RL, Lazar NA. (2016). The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 70 (2): 129-133.
154. Pons J, Adam P. (2018). Biomedical research at the crossroads and ways through. *Medicina Clinica (Barcelona)*. 151 (3): 109-110.

ANNEX A. Exploratory analysis with NC collected in 2014 for the ET study

DESCRIPTIVE RESULTS

Sample of 92 papers published between 24 January 2009 and 3 April 2010:

- 51 randomly assigned to statistical review based on reporting guidelines (SC group)
- 41 randomly assigned to conventional peer review (R group)

This exploratory analysis shows the description of the NC and the Impact Factor sum (IF sum) collected in 2014 for the ET study. We assess the variable IF sum as the sum of the Impact Factors of the journals that publish the citations of the corresponding article in the year of that publication.

The Table A.1 presents the summary statistics of the main variables collected. The mean (sd) follow-up of the articles is 55.38 (3.61) months. The articles of the ET study obtain a mean (sd) of 2.89 (2.81) citations with a mean (sd) of 5.79 (6.65) IF sum.

	n	Mean	SD	Min	Q1	Median	Q3	Max
Number of citations	92	2.89	2.81	0.00	1.00	2.00	4.00	12.00
References	92	21.28	11.49	7.00	10.00	25.50	30.00	57.00
Impact Factor Sum	92	5.79	6.65	0.00	1.03	3.17	8.81	31.78
Months published	92	55.38	3.61	48.63	53.42	55.33	58.43	63.10

Table A.1. Summary statistics of the variables collected in 2014

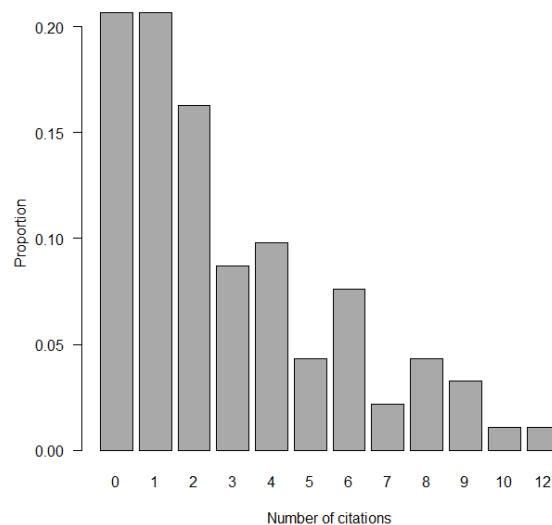


Figure A.1. Proportion of citations of the articles in the ET study. NC collected in 2014

Figure A.2 shows scatter plots between combinations of the previous four variables. Only a relevant relationship arises between the sum of impact factors and the number of citations.

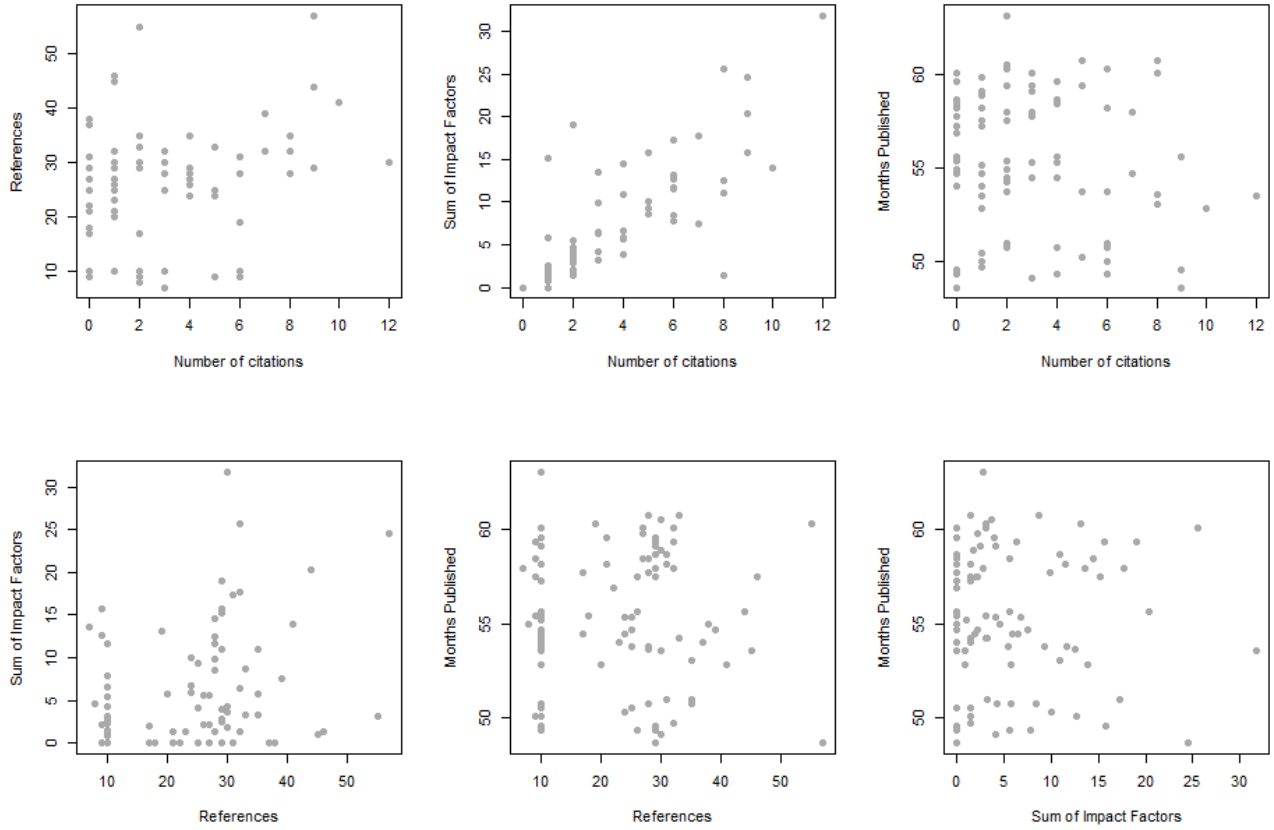


Figure A.2. Scatter plots for the combination of the variables collected in 2014

The next figure shows the relationship between the previous or baseline quality measurements collected for ET study and the two outcome impact variables (NC and IF sum).

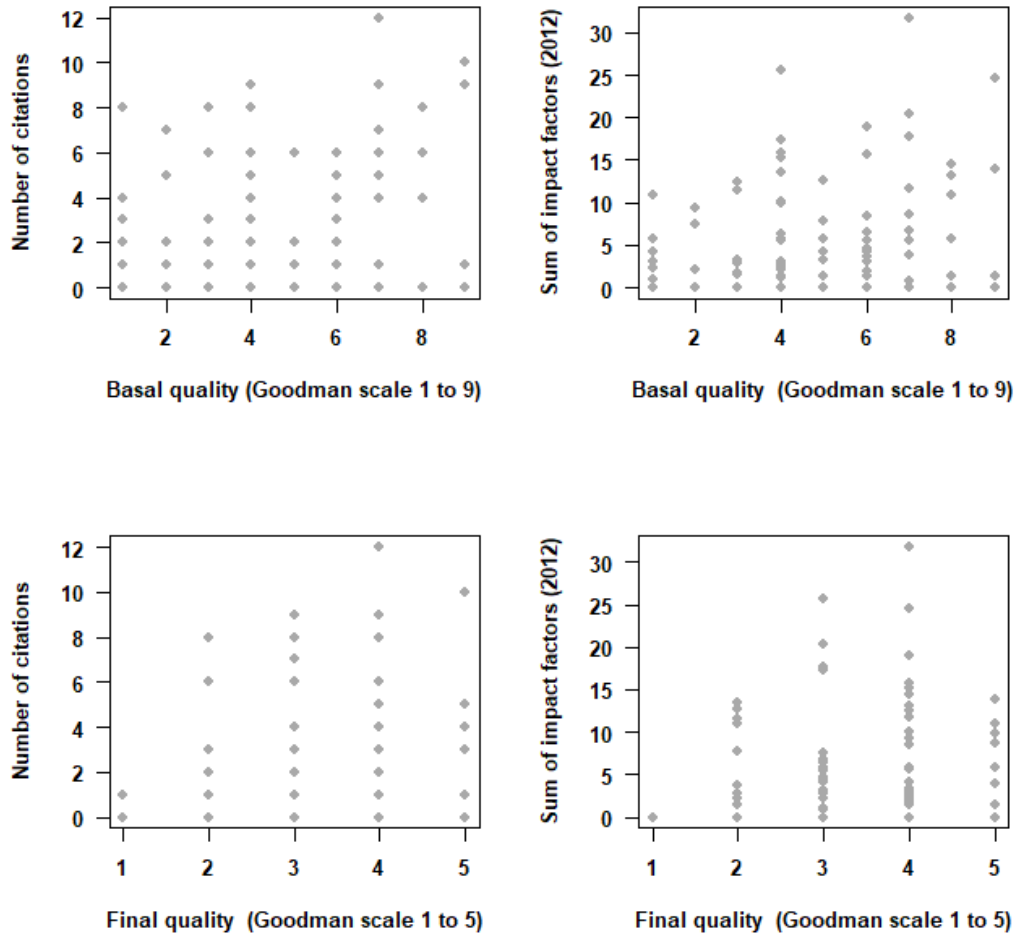


Figure A.3. Relationship between the baseline quality measurements and the NC and the IF sum collected in 2014

In the next figure, regression lines are plotted for all cases (first row), stratified by allocation group (second row) and stratified by study design (third row). The highest slope was found for the intervention studies when taking as response the number of citations, with 0.88 (95% CI: 0.05 to 1.72), as well as the sum of the impact factors, with 1.91 (95% CI: 0.15 to 3.67).

By subgroups

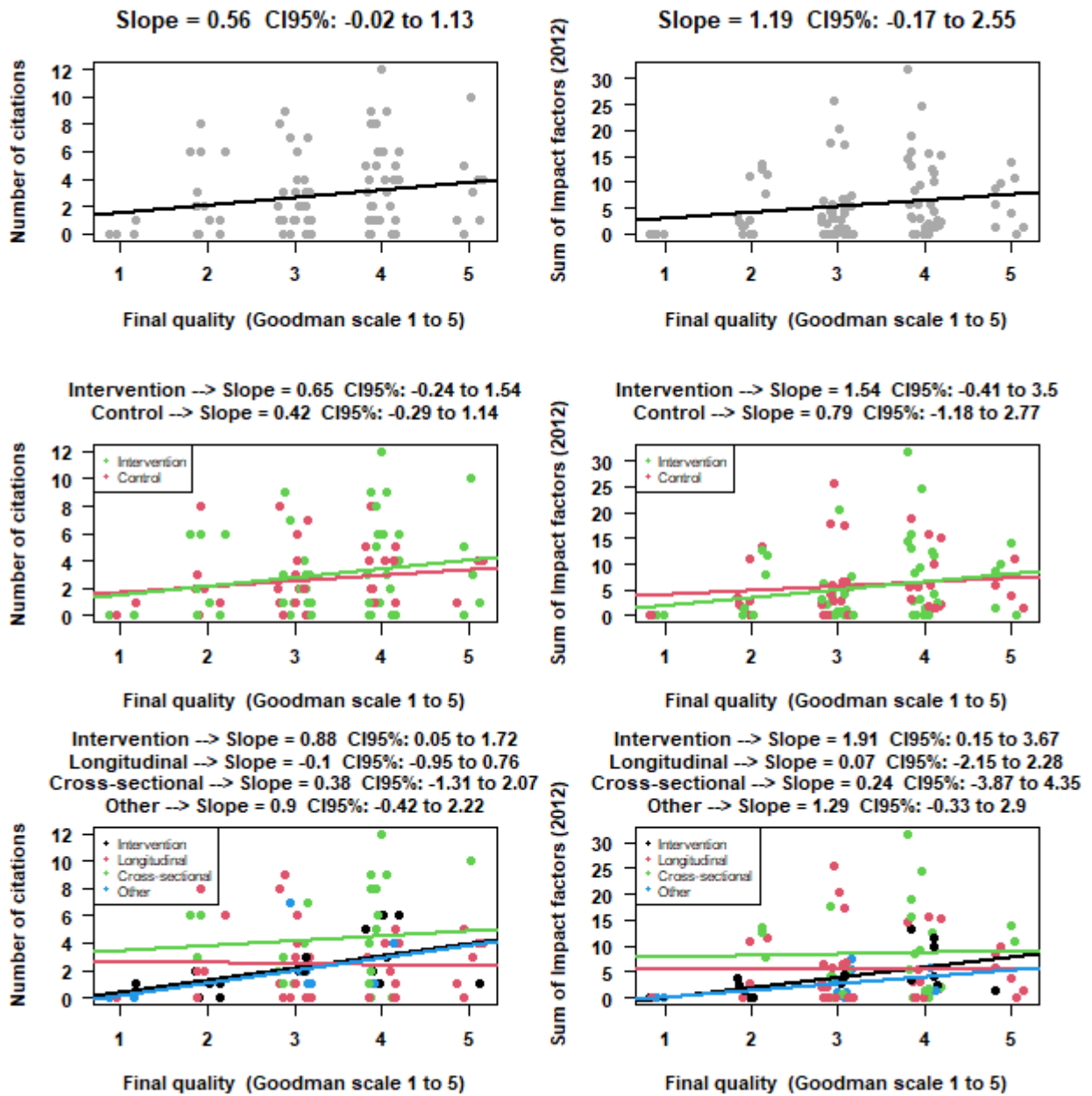
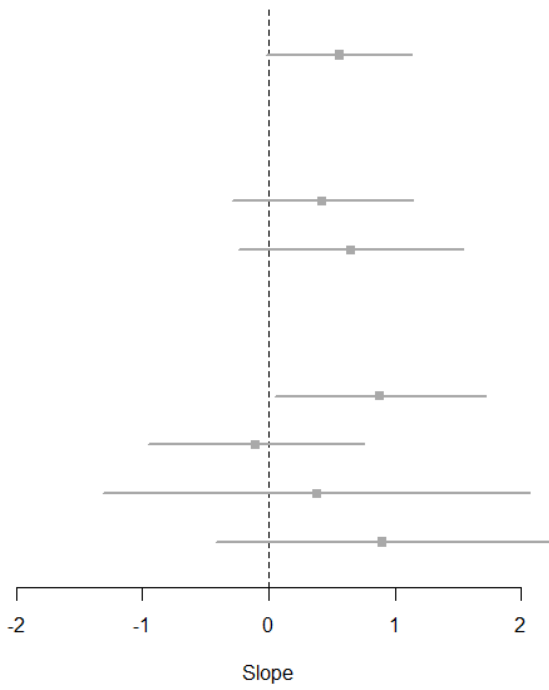


Figure A.4. Regression lines between the final MQAI and the NC and the IF sum collected in 2014, for all cases, by allocation group and by type of study design

95%CI of NC stratified by allocation and study type, and Global



95%CI of sum of IFs stratified by allocation and study type, and Global

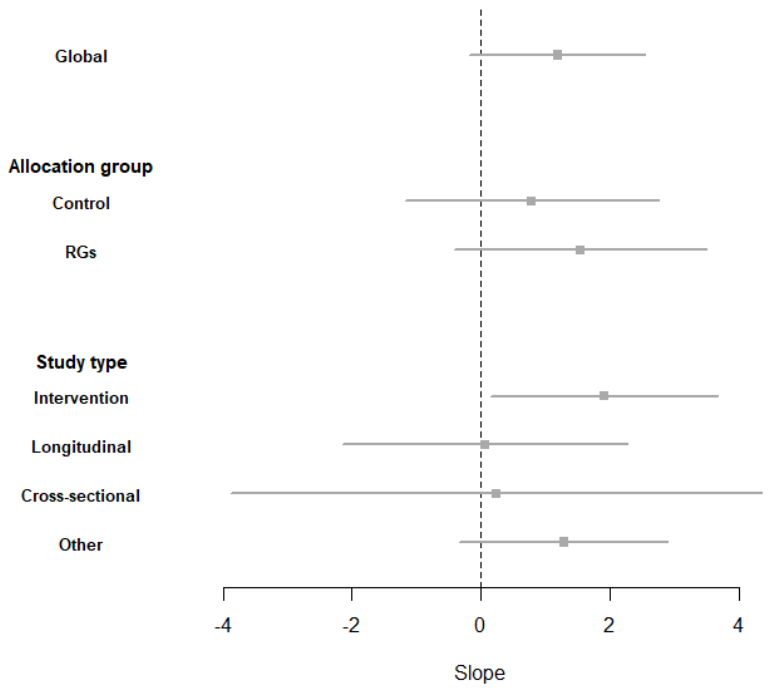


Figure A.5. Forest plot of the slopes of the regression lines between the final MQAI and the NC and the IF sum collected in 2014, for all cases, by allocation group and by type of study design

**DIFFERENCE BETWEEN REFERENCE (R) AND INTERVENTION (SC) GROUPS. –
OUTCOME: NUMBER OF CITATIONS**

Frequencies

	Proportion of citations (row %)											
	0	1	2	3	4	5	6	7	8	9	10	12
- Interventio group (SC)	25.49	17.65	11.76	11.76	3.92	3.92	11.76	1.96	1.96	5.88	1.96	1.96
- Reference group (R)	14.63	24.39	21.95	4.88	17.07	4.88	2.44	2.44	7.32	0.00	0.00	0.00

Table A.2. Proportion of the number of citations by allocation group

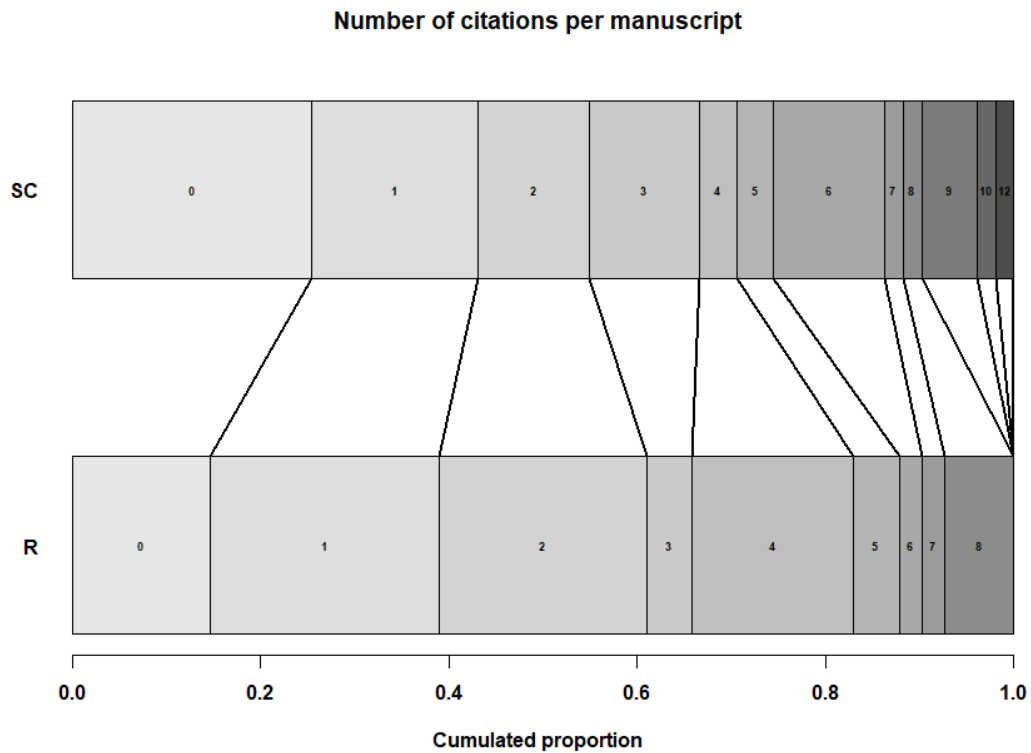


Figure A.6. Cumulative proportion of citations by allocation group

	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	51	3.08	3.17	0.00	0.50	2.00	5.50	12.00
Reference group (R)	41	2.66	2.31	0.00	1.00	2.00	4.00	8.00

Table A.3. Summary statistics of the number of citations by allocation group

**DIFFERENCE BETWEEN REFERENCE (R) AND INTERVENTION (SC) GROUPS. –
OUTCOME: SUM OF IMPACT FACTORS**

	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	51	5.70	6.99	0.00	0.00	3.11	8.96	31.78
Reference group (R)	41	5.91	6.30	0.00	1.40	3.69	6.71	25.60

Table A.4. Summary statistics of the IF sum by allocation group

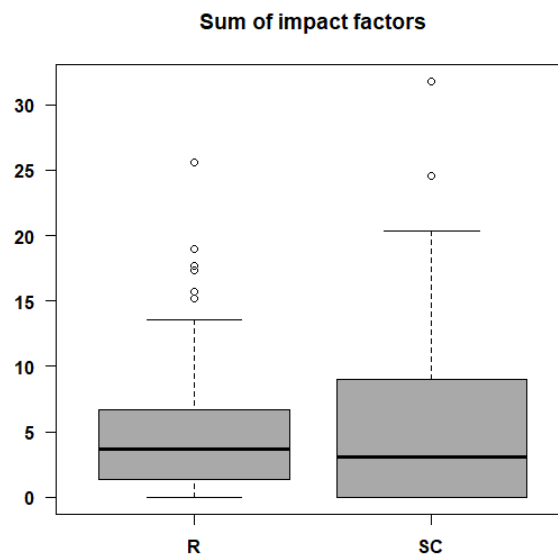


Figure A.7. Boxplot of the IF sum by allocation group

DIFFERENCE BETWEEN REFERENCE (R) AND INTERVENTION (SC) GROUPS BY STUDY DESIGN. –OUTCOME: NUMBER OF CITATIONS

Intervention								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	8	2.63	2.32	0.00	1.00	2.00	3.75	6.00
Reference group (R)	8	1.88	1.46	0.00	1.00	2.00	2.00	5.00
Longitudinal								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	18	2.28	2.52	0.00	0.00	2.00	3.00	9.00
Reference group (R)	20	2.65	2.62	0.00	0.75	1.50	4.00	8.00
Cross-sectional								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	17	4.82	3.84	0.00	1.00	5.00	8.00	12.00
Reference group (R)	9	3.67	2.40	1.00	2.00	3.00	4.00	8.00
Other								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	8	1.63	2.33	0.00	0.00	1.00	2.00	7.0
Reference group (R)	4	2.00	1.41	1.00	1.00	1.50	2.50	4.00

Table A.5. Summary statistics of the number of citations by allocation group and type of study design

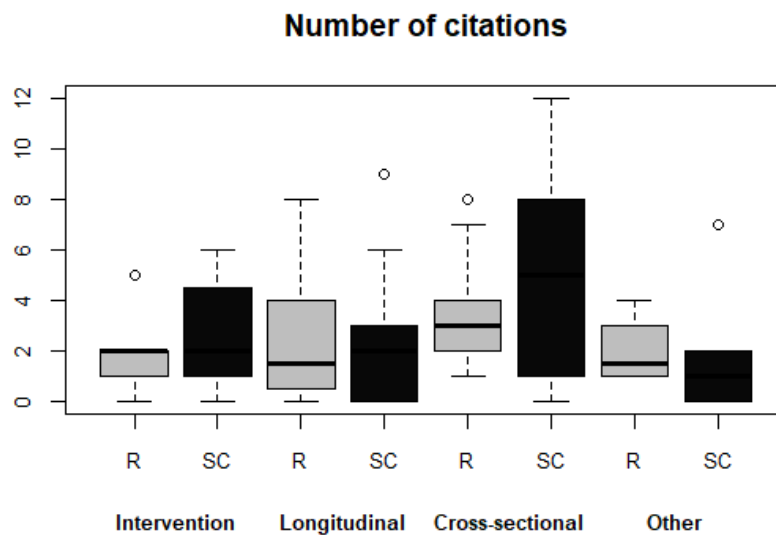


Figure A.8. Boxplot of the number of citations by type of study design

DIFFERENCE BETWEEN REFERENCE (R) AND INTERVENTION (SC) GROUPS BY STUDY DESIGN. –OUTCOME: SUM OF IMPACT FACTORS FOR 2012

Intervention								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	8	4.82	4.91	0.00	1.45	3.33	6.07	13.15
Reference group (R)	8	3.33	3.17	0.00	1.70	3.04	3.92	10.04
Longitudinal								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	18	4.67	6.00	0.00	0.00	2.57	8.08	20.36
Reference group (R)	20	6.37	7.12	0.00	1.05	4.76	7.78	25.60
Cross-sectional								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	17	8.83	9.08	0.00	1.40	7.80	12.63	31.78
Reference group (R)	9	8.50	6.95	1.40	2.16	5.64	13.54	18.99
Other								
	n	Mean	SD	Min	Q1	Median	Q3	Max
Interventio group (SC)	8	2.26	2.86	0.00	0.00	1.00	3.70	7.51
Reference group (R)	4	2.90	2.07	1.40	1.40	2.21	3.71	5.78

Table A.6. Summary statistics of the IF sum by allocation group and type of study design

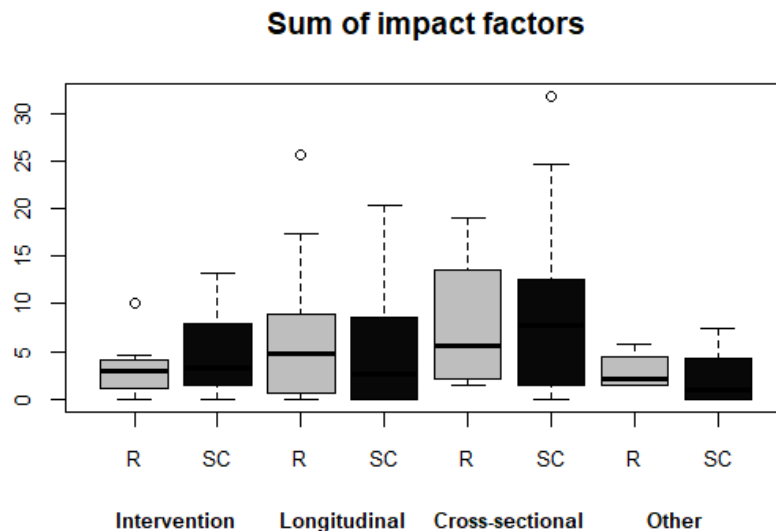


Figure A.9. Boxplot of the IF sum by type of study design

COMPARISON BETWEEN GROUPS (ITT)

Number of citations

Comparison of variances. Original variables

F test to compare two variances

```
data: resp by gr.inv
F = 2.1789, num df = 50, denom df = 40, p-value = 0.01226
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 1.189126 3.913965
sample estimates:
ratio of variances
      2.178934
```

log-transform variables

F test to compare two variances

```
data: log.resp by gr.inv
F = 1.6906, num df = 50, denom df = 40, p-value = 0.08851
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.9226138 3.0367504
sample estimates:
ratio of variances
      1.690582
```

Mean comparison.

Welch Two Sample t-test

```
data: resp by gr
t = -0.9747, df = 87.669, p-value = 0.3324
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -1.6466905 0.5629985
sample estimates:
mean in group 0 mean in group 1
      2.536585      3.078431
```

Sum of impact factor

Two Sample t-test

```
data: resp by gr
t = 0.1443, df = 90, p-value = 0.8856
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.585515 2.990638
sample estimates:
mean in group 0 mean in group 1
      5.905463      5.702902
```

ANNEX B. Assessment of overdispersion for NC

Overdispersion occurs when there is more variability than expected under the response distribution. In Poisson model, the variance should be equal to the mean.

For NC collected in 2018, variance is much larger than mean, indicating overdispersion.

	Mean	Variance	% of zeros
ET study	5.33	24.13	6.52%
IQ study	10.69	236.10	7.84%

Table B.1. Description of the number of citations by study

I explore the zero inflation characteristic by comparing the number of zeros in the data to the number of zeros that expected from a Poisson distribution with a mean equal to the mean count of the data. The proportion of zeros in NC is 6.52% and 7.84% for ET and IQ studies, and the proportion of zeros expected from a Poisson distribution are 1% and 0.002% respectively. Therefore, it seems reasonable to suppose zero inflation for NC.

To explore the distributional properties of the response variable graphically I use a modified version of a histogram that plots the square roots of frequencies rather than the raw frequencies, called rootogram (1).

Figure B.1 shows two hanging rootograms of the NC fitted by a Poisson model for ET and IQ studies. Both rootograms show substantial departures of the model from the data. Specially, the data exhibit too many small counts for a Poisson model to provide an adequate fit.

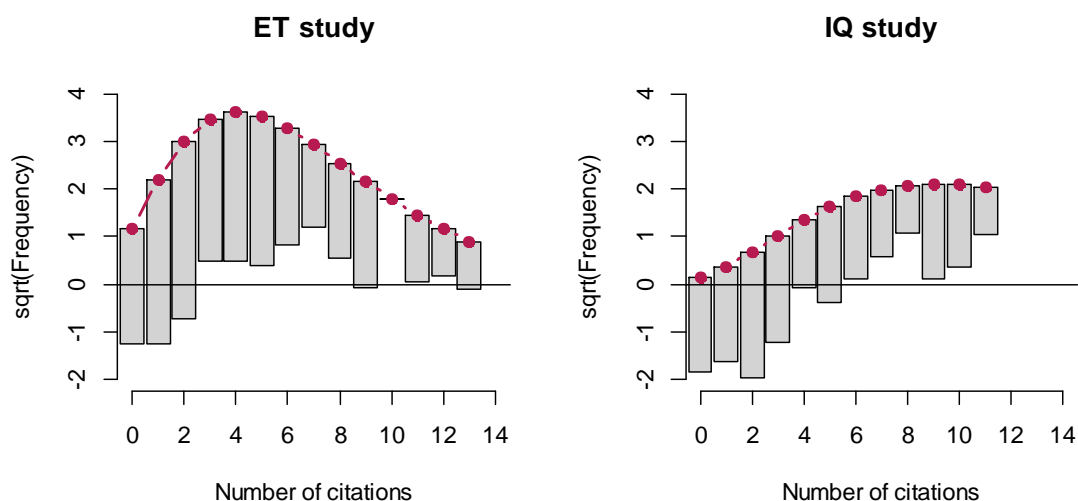


Figure B.1. Rootograms of NC collected in 2018 fitted by Poisson model for ET and IQ studies.

- (1) Kleiber C., Zeileis A. 20016. Visualizing count data regressions using rootograms. *The American Statistician*. 70:3, 296–303.

ANNEX C. Results with NC collected in December 2016 and comparison with 2014 results

DESCRIPTIVE RESULTS

IQ included 99 articles published between 4 February 2005 and 12 May 2006, with a mean (standard deviation, SD) follow-up period of 11.35 (0.31) years. Only the 24 and 27 papers allocated to both interventions and the reference groups respectively were included in the main analysis. The publications received 927 citations (mean 9.36, SD 14.87). ET included 92 articles published between 24 June 2009 and 3 April 2010, with a mean (SD) follow-up period of 7.29 (0.31) years. They received 409 citations (mean 4.44, SD 4.08).

2016		N	Number of citations				Sum NC	Articles with 0 citations	Annual rate Mean (SD)
			Mean	Variance	Median	IQR			
IQ study	R	27	8.4	148.3	4.0	4.0	227	1 (3.7%)	0.74 (1.07)
	S	26	8.4	187.1	4.5	6.75	218	4 (15.4%)	0.74 (1.21)
	C	22	10.3	355.0	4.5	7.5	226	3 (13.6%)	0.90 (1.61)
	SC	24	10.7	241.8	6.5	7.5	256	3 (12.5%)	0.93 (1.34)
	Total	99	9.4	221.0	5.0	7.0	927	11 (11.1%)	0.82 (1.29)
ET study	R	41	3.6	6.3	3.0	2.0	148	2 (4.9%)	0.49 (0.34)
	SC	51	5.1	24.2	3.0	7.0	261	7 (13.7%)	0.72 (0.71)
	Total	92	4.4	16.6	3.0	4.3	409	9 (9.8%)	0.62 (0.58)

Table C.1. Number of citations by study and intervention group in December 2016. Groups not included in main analysis are in a shaded style.

Comparing the number of citations (NC) of 2014 (table B.2.) with the NC of 2016 (table B.1.), we observe an increment in the sum of NC of around 20% for the IQ study and 57% for the ET study. The IQ study increases more citations in the R group (24%), followed by the S (23.9%), C (20.9%) and SC (12.8%) groups. In the ET study intervention group, SC increases by about 66.2% while the R group only by 42.3%. In the IQ study, we see a decrease in the number of articles with 0 citations in those 2 years for only the R and S groups (about 3.7% and 3.8%, respectively); and in the ET study it is about 9.7% for the R and 11.8% for the SC groups.

2014		N	Number of citations				Sum NC	Articles with 0 citations	Annual rate Mean (SD)
			Mean	Variance	Median	IQR			
IQ study	R	27	6.8	94.4	3.0	3.0	183	7.4%	0.72 (1.04)
	S	26	6.8	152.4	3.5	6.0	176	19.2%	0.73 (1.33)
	C	22	8.5	265.3	4.5	6.5	187	13.6%	0.90 (1.69)
	SC	24	7.3	200.4	6.5	7.25	227	12.5%	1.00 (1.46)
	Total	99	7.8	169.2	4.0	5.0	773	13.13%	0.83 (1.37)
ET study	R	41	2.5	4.6	2.0	3.0	104	14.6%	0.48 (0.41)
	SC	51	3.1	10.0	2.0	5.0	157	25.5%	0.60 (0.63)
	Total	92	2.8	7.6	2.0	3.0	261	20.7%	0.54 (0.54)

Table C.2. Number of citations by study and intervention group in December 2014. Groups not included in main analysis are in a shaded style.

Tables C.3 and C.4 show the descriptive results of the number of citations by study, study design and intervention group for the data collected in 2014 and 2016.

For the IQ study:

- *Cross-sectional* is the more frequent design with 0-cited articles, 19% in both 2014 and 2016 collecting times.
- *Longitudinal* is the second design with highest number of 0-cited articles, 14.6% in both 2014 and 2016 collecting times. This design only have citations in intervention group.
- The highest NC mean is for *Other* type of study (13.8 in 2014 and 17.1 in 2016) followed by *Cross-Sectional* (10 in 2014 and 11 in 2016), *Longitudinal* (5.9 in 2014 and 7.6 in 2016) and *Interventional* (3.8 in 2014 and 4.2 in 2016).
- For the variance and median of the NC, we see the same pattern.
- In Cross-sectional studies, the mean NC is the highest for the intervention group SC, with a mean of 24.3 and 25.7 citations for, respectively, 2014 and 2016. As for the variance, it was 862.3 and 904.3, respectively for those same years.

For the ET study:

- *Longitudinal* is the more frequent design with 0-cited articles, (31.6% in 2014 and 18.4 in 2016).
- *Other* is the second design with highest number of 0-cited articles, 25% in 2014 and 8.3% in 2016.
- All study designss have 0 citation counts in the intervention group SC, with the Cross-sectional and Other types having 0 citations only in that group.

- The highest mean and variance values are for Cross-sectional (mean of 4.2 and 6.7, variance of 11.1 and 28.8 for data in 2014 and 2016, respectively) and Longitudinal (mean of 2.5 and 3.8, variance of 6.5 and 12.4 for data in 2014 and 2016, respectively).
- For Cross-sectional and Interventional types of studies, the intervention group SC has a higher mean of NC.

2016			N	Sum NC	Zero (percentage)	NC			
						Mean	Variance	Median	IQR
IQ study	Cross-sectional	R	7	43	14.3	6.1	79.5	4.0	2.5
		S	6	87	33.3	14.5	680.7	4.0	9.0
		C	5	23	20.0	4.6	24.3	3.0	1.0
		SC	3	77	0.0	25.7	904.3	13.0	28.0
		Total	21	230	19.0	11.0	343.7	4.0	10.0
	Interventional	R	3	7	0.0	2.3	1.3	3.0	1.0
		S	4	10	0.0	2.5	3.7	2.0	2.5
		C	5	24	20.0	4.8	17.7	5.0	4.0
		SC	7	39	0.0	5.6	7.3	6.0	2.5
		Total	19	80	5.3	4.2	9.2	3.0	4.5
	Longitudinal	R	12	110	0.0	9.2	164.5	4.0	4.0
		S	12	81	16.7	6.8	69.8	4.5	6.3
		C	7	51	14.3	7.3	91.9	4.0	5.5
		SC	10	68	30.0	6.8	79.7	4.0	8.3
		Total	41	310	14.6	7.6	97.3	4.0	6.0
	Other	R	5	67	0.0	13.4	322.3	6.0	5.0
		S	4	40	0.0	10.0	30.0	8.0	4.5
		C	5	128	0.0	25.6	1296.8	10.0	10.0
		SC	4	72	0.0	18.0	590.0	8.5	15.5
		Total	18	307	0.0	17.1	527.7	9.0	10.0
ET study	Cross-sectional	R	9	37	0.0	4.1	5.1	3.0	1.0
		SC	17	137	5.9	8.1	36.4	8.0	11.0
		Total	26	174	3.8	6.7	28.8	4.0	7.0
	Interventional	R	8	24	0.0	3.0	3.4	2.5	1.3
		SC	8	36	0.0	4.5	7.1	5.0	4.0
		Total	16	60	0.0	3.8	5.5	3.0	3.3
	Longitudinal	R	20	77	10.0	3.9	8.7	4.0	4.3
		SC	18	69	27.8	3.8	17.3	3.0	4.8

		Total	38	146	18.4	3.8	12.4	3.0	4.0
	Other	R	4	10	0.0	2.5	3.0	2.5	3.0
		SC	8	19	12.5	2.4	6.0	2.0	1.3
		Total	12	29	8.3	2.4	4.6	2.0	2.3

Table C.3. Number of citations by study, study design and intervention group in 2016. Groups not included in main analysis are in a shaded style.

2014			N	Sum NC	Zero (percentage)	NC			
						Mean	Variance	Median	IQR
IQ study	Cross-sectional	R	7	43	14.3	6.1	79.5	4.0	2.5
		S	6	87	33.3	14.5	680.7	4.0	9.0
		C	5	23	20.0	4.6	24.3	3.0	1.0
		SC	3	77	0.0	25.7	904.3	13.0	28.0
		Total	21	230	19.0	11.0	343.7	4.0	10.0
	Interventional	R	3	7	0.0	2.3	1.3	3.0	1.0
		S	4	10	0.0	2.5	3.7	2.0	2.5
		C	5	24	20.0	4.8	17.7	5.0	4.0
		SC	7	39	0.0	5.6	7.3	6.0	2.5
		Total	19	80	5.3	4.2	9.2	3.0	4.5
	Longitudinal	R	12	110	0.0	9.2	164.5	4.0	4.0
		S	12	81	16.7	6.8	69.8	4.5	6.3
		C	7	51	14.3	7.3	91.9	4.0	5.5
		SC	10	68	30.0	6.8	79.7	4.0	8.3
		Total	41	310	14.6	7.6	97.3	4.0	6.0
	Other	R	5	67	0.0	13.4	322.3	6.0	5.0
		S	4	40	0.0	10.0	30.0	8.0	4.5
		C	5	128	0.0	25.6	1296.8	10.0	10.0
		SC	4	72	0.0	18.0	590.0	8.5	15.5
		Total	18	307	0.0	17.1	527.7	9.0	10.0
ET study	Cross-sectional	R	9	37	0.0	4.1	5.1	3.0	1.0
		SC	17	137	5.9	8.1	36.4	8.0	11.0
		Total	26	174	3.8	6.7	28.8	4.0	7.0
	Interventional	R	8	24	0.0	3.0	3.4	2.5	1.3
		SC	8	36	0.0	4.5	7.1	5.0	4.0
		Total	16	60	0.0	3.8	5.5	3.0	3.3
	Longitudinal	R	20	77	10.0	3.9	8.7	4.0	4.3

		SC	18	69	27.8	3.8	17.3	3.0	4.8
		Total	38	146	18.4	3.8	12.4	3.0	4.0
	Other	R	4	10	0.0	2.5	3.0	2.5	3.0
		SC	8	19	12.5	2.4	6.0	2.0	1.3
		Total	12	29	8.3	2.4	4.6	2.0	2.3

Table C.4. Number of citations by study, study design and intervention group in 2014. Groups not included in main analysis are in a shaded style.

MAIN COMPARISON

The main comparison of the analysis compares the reference (R) with the additional intervention of an expert statistician who uses the reporting guidelines (SC). The latter is the shared intervention group in the ET and IQ studies.

For the 2016 NC (Figure C.1), the point estimate of the effect is positive for both the ET and IQ studies, indicating that the intervention increases the number of citations. The joint effect of the two studies indicates that the intervention increases the citation-per-year mean ratio by 43% (95% CI: 3% to 98%). Similar results are obtained when considering raw count data and Poisson distribution (through a Generalized Linear Model). The estimated joint effect for non-adjusted and adjusted Poisson model are 40% (95% CI: 2 to 93%) and 50% (95% CI: 6 to 112%), respectively. 95% CIs are calculated using Jackknife method.

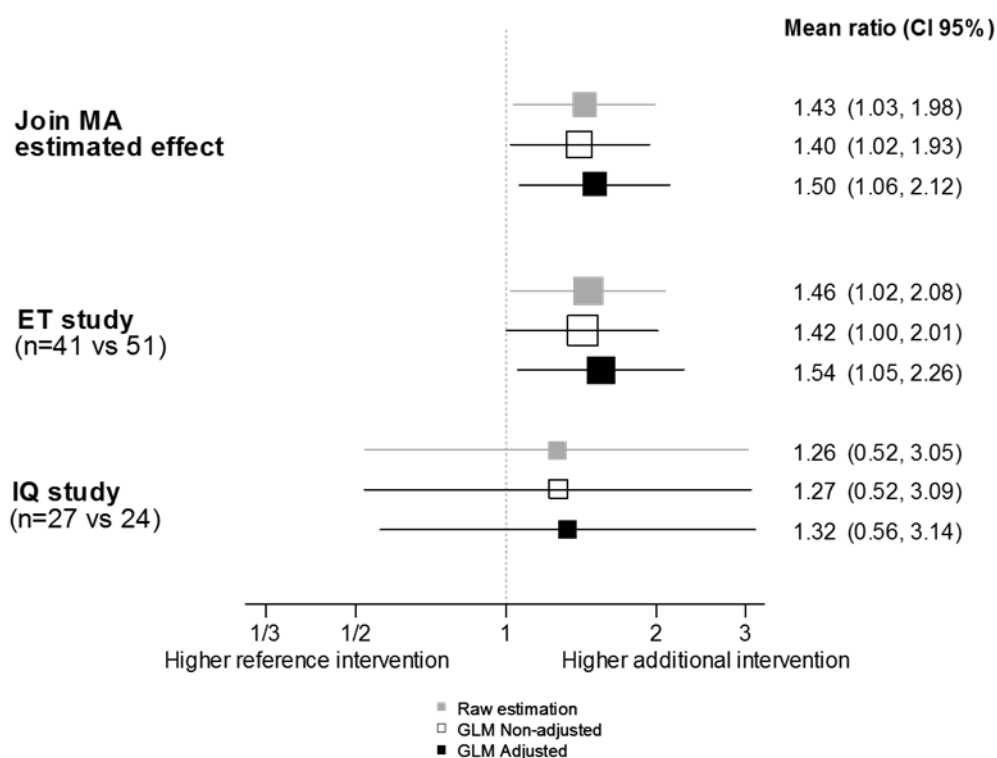


Figure C.1. Citations-per-year mean ratio for the main comparison of NC collected in 2016.

Forest plot in Figure C.1 reflects that the intervention effect is homogenous between the ET and IQ studies. It shows that the confidence intervals are all overlapping and in addition to that, the ET and IQ studies favour the additional intervention.

Figure C.2 shows that the effects in the main comparison are consistent over time. For data collected in 2016, the effect is higher, increasing from 1.26 (95% CI: 0.88 to 1.81) in 2014 to 1.43 (95% CI: 1.03 to 1.98). This consistency makes the effect more plausible, suggesting that the effect of the intervention could apply with shorter follow-up, although sampling error should be reduced, for example examining a larger sample.

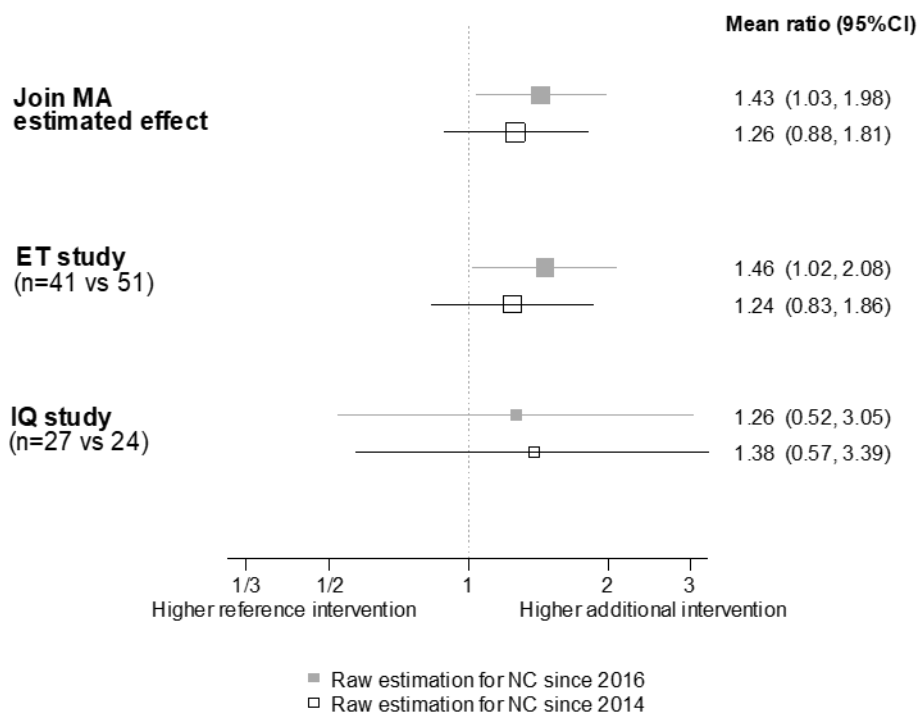


Figure C.2. Citation-per-year mean ratio for the main comparison of NC collected in 2014 and in 2016

SUBGROUP ANALYSIS

We have also explored the effect of the two additional interventions in the IQ study: the statistician (S+SC) and reporting guidelines (C+SC).

Figure C.3 shows the effect of the additional intervention of a statistical reviewer, obtaining a mean-per-year ratio of 1.03 (95% CI: 0.53 to 2.01). For the additional intervention of reporting guidelines, the effect is also positive, with a mean ratio of 1.24 (95% CI: 0.64 to 2.40).

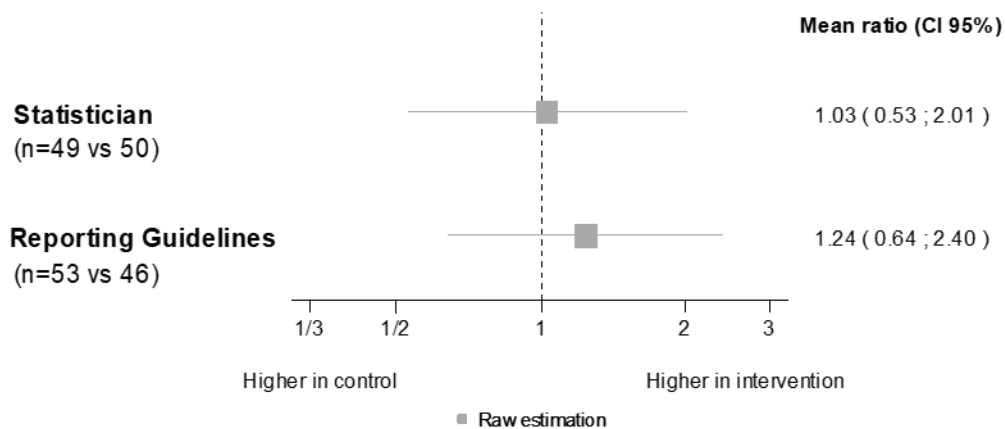


Figure C.3. Citations-per-year mean ratio for Statistician and Reporting Guidelines effects (only the IQ study)

Another subgroup analysis that we studied is the effect of the additional intervention of a statistician using reporting guidelines (SC), which is common in both the ET and IQ studies by study design.

Figure C.4 shows that the estimated joint effect for cross-sectional and interventional study designs increase the mean-per-year ratio by 2.3 (95% CI: 1.37 to 3.86) and 1.87 (95% CI: 1.15 to 3.03), respectively. For the ET study, the type of study with a higher effect is Cross-sectional, with a mean-per-year ratio of 2.13 (95% CI: 1.22 to 3.69). For the IQ study, is the interventional study design with a mean-per-year ratio of 2.44 (95% CI: 1.18 to 5.04). 95% CI are obtained using Jackknife method.

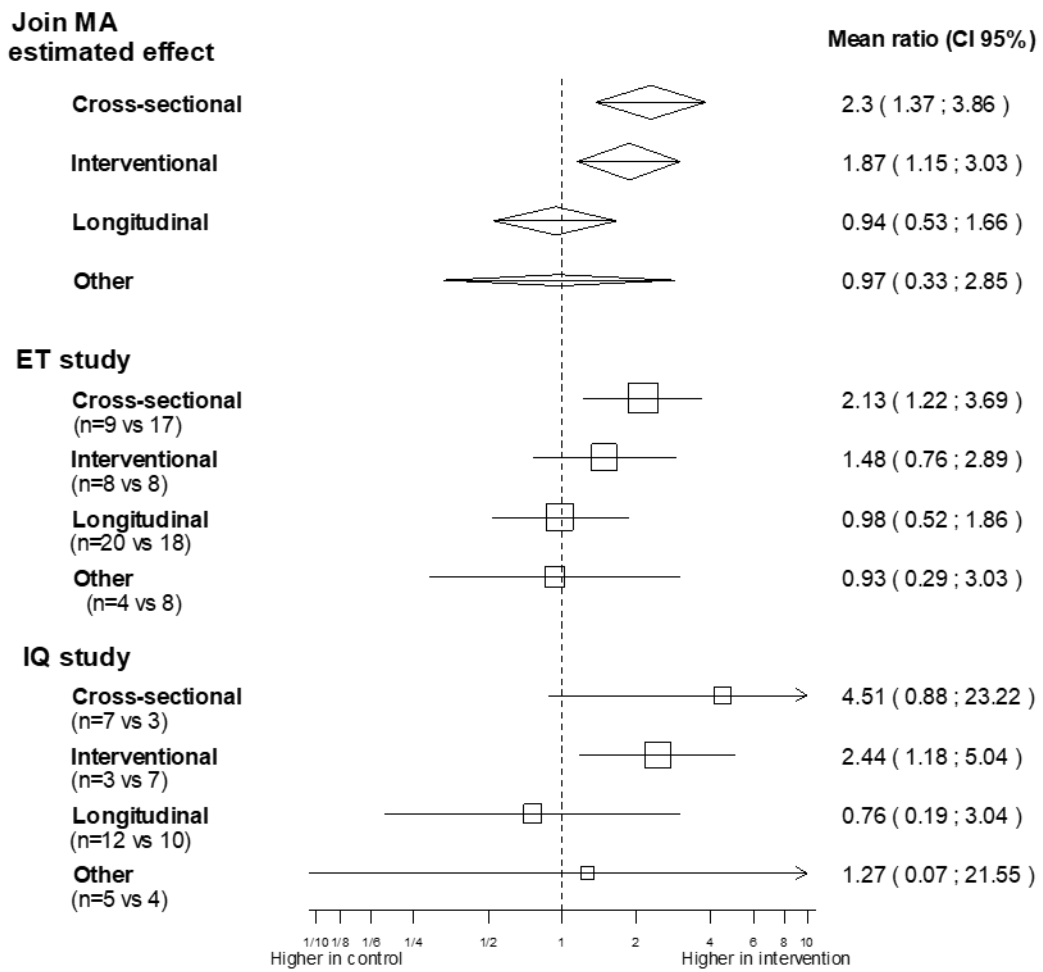


Figure C.4. Citations-per-year mean ratio of the additional intervention of a statistical reviewer using reporting guidelines by study design

ANNEX D. Estimation of the intervention effect on non-citable articles using zero-inflated models

In the first exploratory analysis conducted in 2014 it seemed that the model that best fit the NC took into account overdispersion due to non-citable articles. When we explored the NC with greater follow-up, we realized that dispersion due to a large number of 0-citations was no longer relevant. Figure D.1, that shows the results of the comparison of non-citable articles by group, confirms this. It has very large CI values, indicating that these models do not converge, probably due to the lack of information to estimate this parameter.

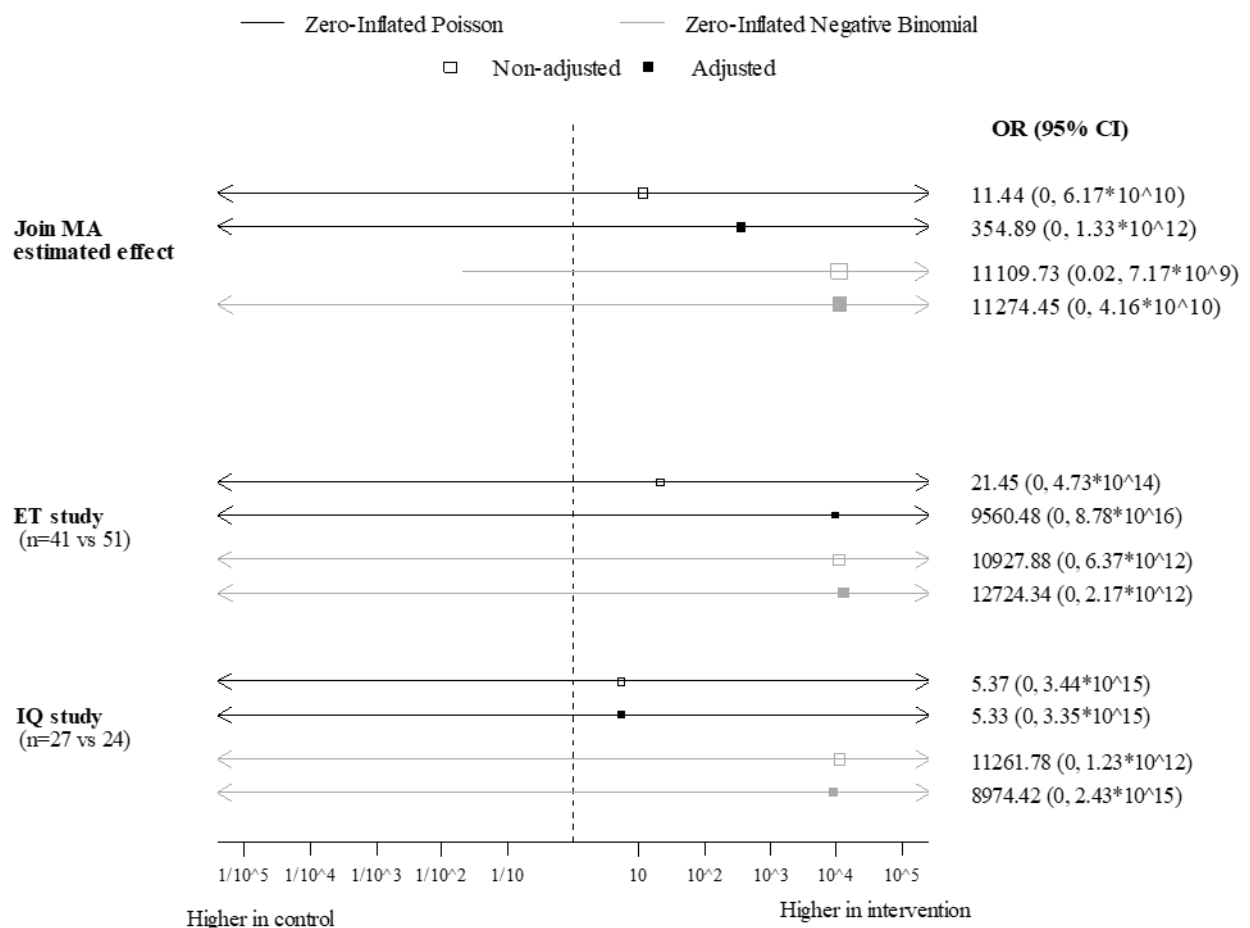


Figure D.1. Odds ratio (OR) estimation for the main comparison (R vs. SC) of non-citable articles using zero inflated models. 95% confidence intervals (CI) are estimated using Jackknife method.

ANNEX E. Comparison of sample density and theoretical densities of Poisson, QP, ZIP, NB and ZINB

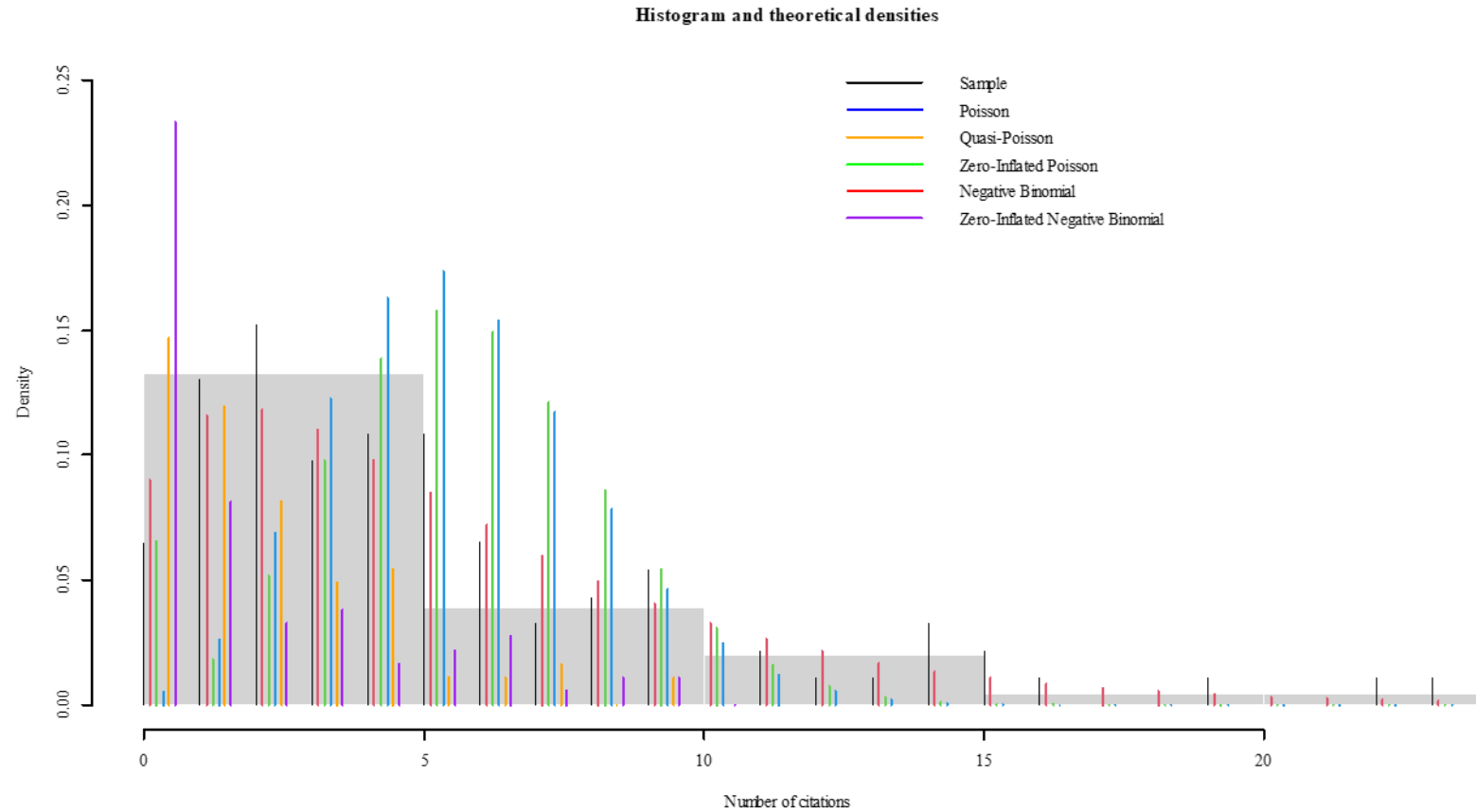


Figure E.1. Histogram of the ET study data with sample density and theoretical densities of Poisson, QP, ZIP, NB and ZINB

ANNEX F. Characteristics of recurrent events data models

	Risk interval	Risk set for citation k at time t	Baseline hazard	Within-article correlation	Comment
AG	Duration since article publication	Independent citations (any given citation is not affected by previous citations)	Common baseline hazard across all citations of all articles	The within article citations are independent	AG models is recommended when there is no citation dependence and no covariate citation effects
PWP-TT	Duration since article publication	All articles that have experienced citation k-1, and have not experienced citation k at time t	Stratifies the data by citation count so that the baseline hazard is allowed to vary with each citation	The within-article citations are independent	PWP-TT model takes into account the ordering of events
PWP-GT	Duration since previous citation	All articles that have experienced citation k-1, and have not experienced citation k at time t	Stratifies the data by citation count so that the baseline hazard is allowed to vary with each citation	The within-article citations are independent	PWP-GT model takes into account the ordering of events
WLW	Duration since article publication	All articles that have not experienced citation k at time t	Stratifies the data by citation count so that the baseline hazard is allowed to vary with each citation	The within-article citations are independent	At any time point, WLW-TT describes all articles that have not yet experienced k citations and are assumed to be at risk for the k th citation
Frailty	Duration since article publication	A random effect (frailty) term is used to account for the within-article correlation between citations to enable modeling the phenomenon by which some articles are intrinsically more or less susceptible to receive citations than others	Heterogeneity is directly incorporated via a random effect so that the baseline hazard is allowed to vary with each citation	Captures within-article correlation due to both citation dependence and heterogeneity	The frailty approach accounts for heterogeneity. The random effect (the frailty) has a multiplicative effect on the baseline hazard function and the mixture of articles with different citation risks

Table F.1. Characteristics of recurrent events data models

ANNEX G. Simulation study to compare shared frailty models for recurrent events

In order to perform simulations, we need to generate recurrent event data. The occurrences times from shared frailty models assume independence and follow a proportional hazards model, given the frailty (1).

The survival function for article i in citation time j with z_i frailty parameter is

$$S_{ij}(t) = \exp(-H_0(t)z_i \exp(x_{ij}\beta)) \quad [D1]$$

Then, the cumulative distribution function of the proportional hazards model is

$$F_{ij}(t) = 1 - \exp(-H_0(t)z_i \exp(x_{ij}\beta)) \quad [D2]$$

The distribution function $F(t)$, denoted by V , follows a uniform distribution on the interval 0 to 1.

If $V \sim U[0,1]$, then $(1 - V) \sim U[0,1]$. Thus, let t be the survival time of the proportional hazards model, then it follows from [D.2] that

$$V = \exp(-H_0(t)z_i \exp(x_{ij}\beta)) \sim U[0,1] \quad [D3]$$

If $h_0(t) > 0$ for all t , then H_0 can be inverted and the survival time t can be expressed as

$$t = H_0^{-1}\left(-\frac{\log(V)}{z_i \exp(x_{ij}\beta)}\right) \quad [D4]$$

Where H_0 is the cumulative baseline hazard function.

When considering that the baseline hazard function H_0 follows an exponential distribution $H_0^{-1}(\cdot) = \frac{t}{\lambda}$, the occurrence time is

$$t = \frac{\left(-\frac{\log(V)}{z_i \exp(x_{ij}\beta)}\right)}{\lambda} \quad [D5]$$

For simplicity, the frailty term z_i is generated from a gamma distribution.

The intervention effect parameter β is set to $\ln(1.5)$ so that the hazard rate of articles in the intervention group is 1.5 times the hazard rate in the reference group.

We generate $n_{sim}=1000$ data sets under the exponential baseline hazard model and gamma frailty model. Using the different packages available in R software for shared frailty models we estimate the parameters for the variance of the frailty distribution, θ and the fixed treatment effect β . We also compute the bias and mean square error (MSE) of the estimates following these formulas:

$$Bias(\theta) = \frac{\sum_{r=1}^{nsim} \hat{\theta}_r}{nsim} - \theta \quad [D6]$$

$$MSE(\theta) = Bias^2(\theta) + \frac{\sum_{r=1}^{nsim} Var(\hat{\theta}_r)}{nsim} \quad [D7]$$

$$Bias(\beta) = \frac{\sum_{r=1}^{nsim} \hat{\beta}_r}{nsim} - \beta \quad [D8]$$

$$MSE(\beta) = Bias^2(\beta) + \frac{\sum_{r=1}^{nsim} Var(\hat{\beta}_r)}{nsim} \quad [D9]$$

Where $\hat{\theta}_r$ and $\hat{\beta}_r$ are the estimates of interest within each of the $nsim$ simulations.

We assume that the frailty term follows a gamma distribution with the shape parameter 2 and scale parameter 0.5. The mean and variance of the gamma frailty distribution are 1 and 0.5, respectively.

Table G.1. shows the bias and MSE for θ and β , estimated for the R packages that allow the frailty term to follow a gamma distribution. The `parfm` function of the `parfm` package does not converge and the `coxph` function does not estimate the variance of θ needed for the estimation of the MSE. The bias for β are similar for survival and `frailtyEM` packages. With the `frailtypack` package the bias and MSE for θ parameter are the lowest. There is no evidence for choosing one among the others.

package::function	Bias(β)	MSE(β)	Bias(θ)	MSE(θ)
<code>parfm::parfm</code>	nc	nc	nc	nc
<code>survival::coxph</code>	0.0066	0.0307	-0.1262	*
<code>frailtypack::frailtyPenal</code>	0.0133	0.0308	-0.0112	0.0091
<code>frailtyEM::emfrail</code>	0.0066	0.0307	0.1134	0.0434

* `survival::coxph` do not show the variance of θ

Table G.1. Bias and MSE of the variance of the frailty distribution, θ and the fixed treatment effect β , obtained with the simulation study for different R packages. *nc=no converge*.

The estimation method used in `frailtypack` is the maximization of the penalized likelihood, which is described in detail in Rondeau et al (2).

- (1) Bender R., Augustin T, and Blettner M. 2005. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* 24: 1713–1723.
- (2) Rondeau V, Mazroui Y, Gonzalez JR. 2012. `frailtypack`: An R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation.

R code for the simulation study

```

rm(list=ls())
set.seed(1234)
library(parfm)
library(frailtypack)
library(frailtyEM)
library(survival)

num.sim <- 1000
#- Cox (frailty)
estim.cox.frailty <- matrix(NA, ncol=6, nrow=num.sim)
names(estim.cox.frailty) <-
c("effect", "effect.inf", "effect.sup", "theta.estim", "theta.estim.inf", "theta.estim.sup")
#- EM
estim.frailtyEM <- matrix(NA, ncol=6, nrow=num.sim)
names(estim.frailtyEM) <-
c("effect", "effect.inf", "effect.sup", "theta.estim", "theta.estim.inf", "theta.estim.sup")
#- FrailtyPenal Gamma
estim.frailtyPenal <- matrix(NA, ncol=6, nrow=num.sim)
names(estim.frailtyPenal) <-
c("effect", "effect.inf", "effect.sup", "theta.estim", "theta.estim.inf", "theta.estim.sup")

for (i in 1:num.sim){
  Nart = 90
  NT = 10 # años
  theta = 2
  frail = rgamma(Nart, shape=theta, scale=1/theta)
  beta = log(1.5)
  arm = sample(0:1, Nart, replace=TRUE)
  lambda = 1 # Weibull scale parameter
  rho = 1 # Weibull shape parameter (rho=1 if exponential(1/lambda))

  Y.start = c(); Y.stop = c(); Z = c(); Ev = c(); G = c()

  for (a in 1:Nart) {
    v <- runif(n=100)
    x <- (- log(v) / (lambda * frail[a] * exp(arm[a] * beta)))^(1 / rho)
    xx = cumsum(x)
    y.start = c(0, xx[xx<NT])
    y.stop = c(xx[xx<NT], NT)
    z = rep(a, length(y.start))
    Y.start = c(Y.start, y.start)
    Y.stop = c(Y.stop, y.stop)
    Z = c(Z, z)
    Ev = c(Ev, c(rep(1, length(y.start)-1), 0))
    G = c(G, rep(arm[a], length(y.start)))
  }

  dat.sim <- data.frame(cbind(Y.start, Y.stop, Z, Ev, G))
  dat.sim$Ev <- as.numeric(as.character(dat.sim$Ev))
  dat.sim$Y.start <- as.numeric(as.character(dat.sim$Y.start))
  dat.sim$Y.stop <- as.numeric(as.character(dat.sim$Y.stop))

  #- Cox model (Frailty)
  fit3 <- coxph(Surv(Y.start, Y.stop, Ev) ~ G + frailty(Z, distribution="gamma"), dat = dat.sim)
  effect <- exp(fit3$coefficients)
  effect.inf <- exp(fit3$coefficients - 1.96 * sqrt(fit3$var))
  effect.sup <- exp(fit3$coefficients + 1.96 * sqrt(fit3$var))
  theta.estim <- 1 / as.numeric(substr(summary(fit3)$print2, start=28, stop=36))
  theta.estim.inf <- ""
  theta.estim.sup <- ""
  estim.cox.frailty[i, ] <-
c(effect, effect.inf, effect.sup, theta.estim, theta.estim.inf, theta.estim.sup)

  #- EM
  fit1 <- emfrail(Surv(Y.start, Y.stop, Ev) ~ G + cluster(Z), dat = dat.sim)
  effect <- exp(fit1$coefficients)
  effect.inf <- exp(fit1$coefficients - 1.96 * summary(fit1)$coef[, 3])
  effect.sup <- exp(fit1$coefficients + 1.96 * summary(fit1)$coef[, 3])
  theta.estim <- exp(fit1$logtheta)
  theta.estim.inf <- exp(fit1$ci_logtheta[1])
  theta.estim.sup <- exp(fit1$ci_logtheta[2])
}

```

```
estim.frailtyEM[i,] <- c(effect, effect.inf, effect.sup, theta.estim, theta.estim.inf, theta.estim.sup)

#- FrailtyPenal Gamma
fit6 <- frailtyPenal(Surv(Y.start, Y.stop, Ev)~G + cluster(Z), data=dat.sim, n.knots=7, kappa=10000)
effect <- exp(fit6$coef)
effect.inf <- exp(fit6$coef-1.96*sqrt(fit6$varHIH))
effect.sup <- exp(fit6$coef+1.96*sqrt(fit6$varHIH))
theta.estim <- fit6$theta
theta.estim.inf <- fit6$theta-1.96*sqrt(fit6$varTheta)[1]
theta.estim.sup <- fit6$theta+1.96*sqrt(fit6$varTheta)[1]
estim.frailtyPenal.gamma[i,] <-
c(effect, effect.inf, effect.sup, theta.estim, theta.estim.inf, theta.estim.sup)
```

ANNEX H. Results of the adjusted Frailty models for recurrent events

This annex complements the analysis of the NC as recurrent events with a semi-parametric frailty model with the estimations obtained adjusting for the baseline MQAI value of the articles.

For the ET study, the hazard of receiving a citation for an article in the SC group at any time t compared to an R group article is 1.41 (95% CI: 0.99 to 2.01) according to the survival package. The estimation for the frailtyEM package is similar with a HR equals to 1.41 (95% CI: 0.99 to 2.00). The estimation for the frailtypack package is 1.36 (95% CI: 1.03 to 1.81). For all packages the estimations suggest that the articles in the SC group have higher hazard to receive a citation.

For the IQ study, the hazard of receiving a citation for an article in the SC group at any time t compared to an R group article is 1.16 (95% CI: 0.36 to 3.67) according to the survival package. The estimation for the frailtyEM package is similar with a HR equals to 1.16 (95% CI: 0.62 to 2.19). The estimation for the frailtypack package is 1.29 (95% CI: 0.79 to 2.11). For all packages the estimations suggest that the intervention has positive effect on the NC of the articles.

<i>Adjusted analysis</i>	ET study		IQ study	
	HR (95% CI)	Variance of frailty (95% CI)	HR (95% CI)	Variance of frailty (95% CI)
survival::coxph	1.41 (0.99, 2.01)	0.53 (-)	1.16 (0.36, 3.67)	1.21 (-)
frailtyEM::emfrail	1.41 (0.99, 2.00)	0.52 (0.34, 0.78)	1.16 (0.62, 2.19)	1.19 (0.80, 1.81)
frailtypack::frailtyPenal	1.36 (1.03, 1.81)	0.25 (0.10, 0.41)	1.29 (0.79, 2.11)	1.23 (0.75, 2.03)

Table H.1. Estimation of the hazard ratio between the NC for R and SC groups adjusted for the baseline MQAI for ET and IQ studies.

ANNEX I. Main results with NC collected in July 2020

Figure I.1 shows the raw, the Poisson non-adjusted and the Poisson adjusted estimates of the SC additional intervention using the mean citations-per-year ratio for the NC collected in July 2020. All reported 95% CI come from the JK method.

For the ET study, the intervention increases the mean citations-per-year by 32% (95% CI: -6 to 87%) using raw estimation, by 40% (95% CI: -3 to 103%) using Poisson non-adjusted and by 38% (95% CI: -4 to 99%) using Poisson adjusted.

For the IQ study, the intervention increases the mean citations-per-year by 15% (95% CI: -52 to 173%) using raw estimation, by 17% (95% CI: -50 to 177%) using Poisson non-adjusted and by 17% (95% CI: -50 to 177%) using Poisson adjusted.

The joint MA estimated effect is 29% (95% CI: -6 to 78%) for the raw estimation, 36% (95% CI: -3 to 90%) for Poisson non-adjusted and 34% (95% CI: -4 to 87%) for Poisson adjusted.

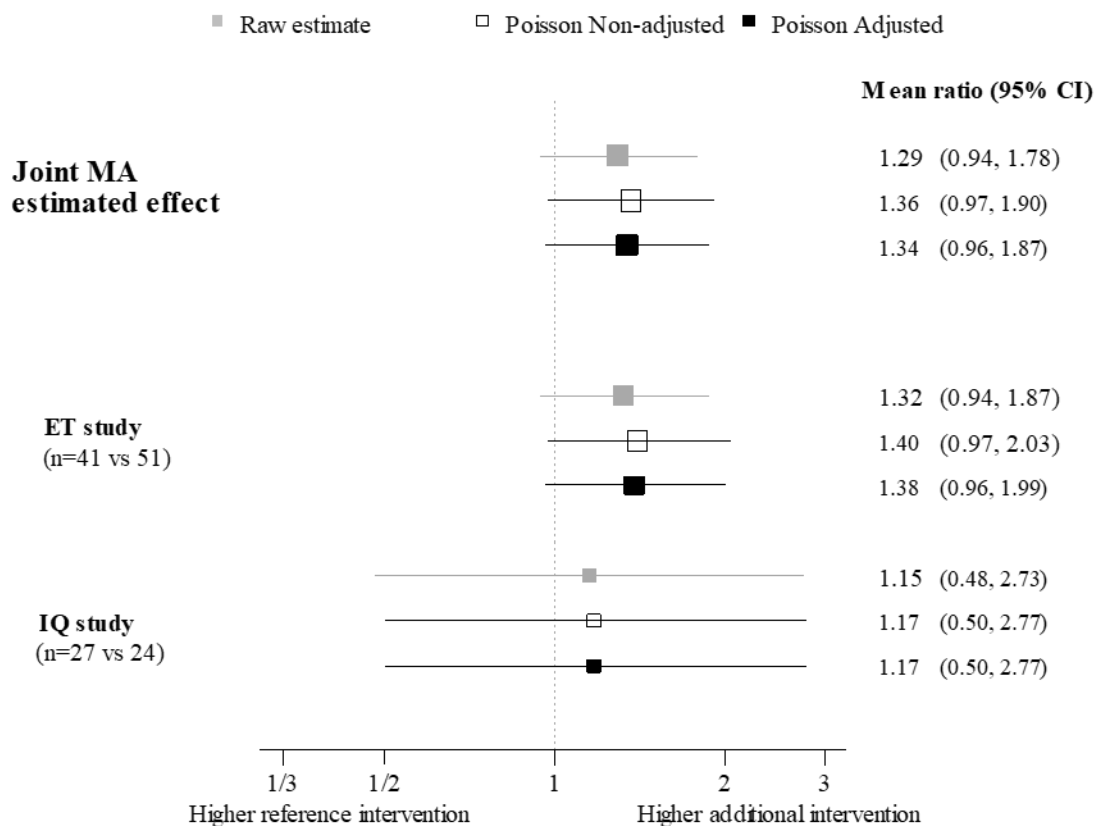


Figure I.1. Citations-per-year mean ratio for the main comparison (R vs SC) with NC collected in July 2020

Figure I.2. shows the MCF for the reference and the intervention groups with a follow-up of more than 11 years. This figure shows the same as Figure 19, the MCF is greater for the SC than for the R group and that the cumulative NC growth seems constant in the SC group, while this growth decrease with time for the R group, mainly after four years after publication.

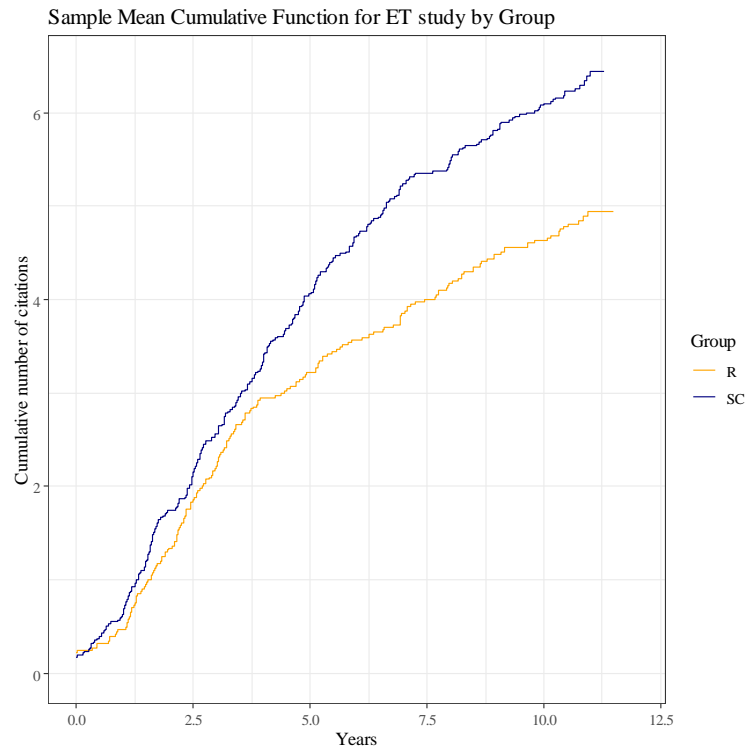


Figure I.2. Mean cumulative function for the ET study by group with NC collected in July 2020

Figure I.3 presents the MCF differences between the intervention and reference groups, with their random uncertainty measured as 95% CI. This figure reflects that the MCF difference stabilize after the seventh year after publication. This difference is always in favor of the intervention group, persisting a positive effect of the intervention.

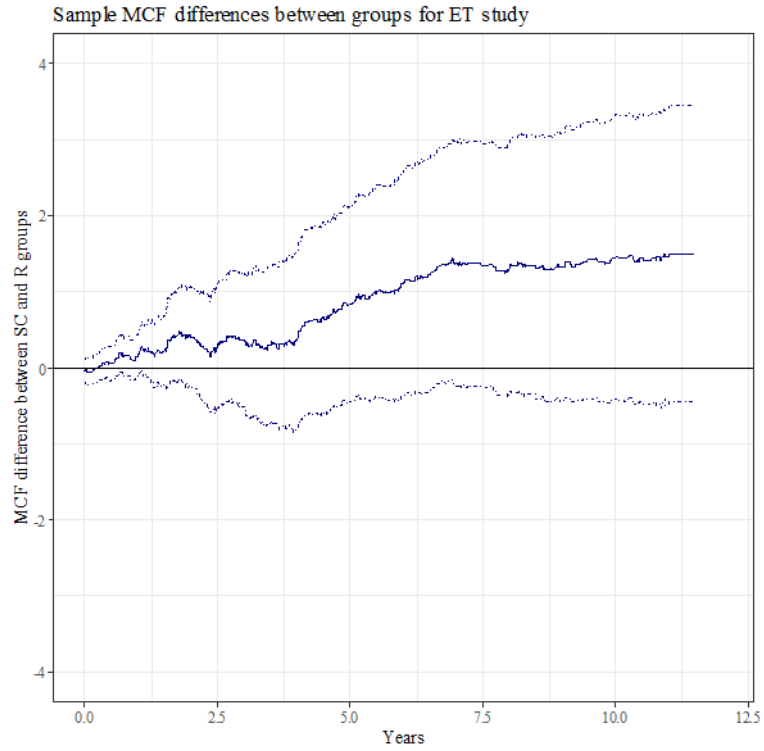


Figure I.3. Mean cumulative function differences between groups for the ET study with NC collected in 2020

ANNEX J. Article published in BMC Medical Research Methodology

Vilaró et al. *BMC Medical Research Methodology* (2019) 19:112
<https://doi.org/10.1186/s12874-019-0746-4>

BMC Medical Research
Methodology

RESEARCH ARTICLE

Open Access

Adherence to reporting guidelines increases the number of citations: the argument for including a methodologist in the editorial process and peer-review



Marta Vilaró^{1,12*} , Jordi Cortés¹, Albert Selva-O'Callaghan^{2,3,4}, Agustín Urrutia^{2,3,5}, Josep-Maria Ribera^{2,3,6}, Francesc Cardellach^{2,7}, Xavier Basagaña^{8,9,10}, Matthew Elmore¹, Miquel Vilardell^{2,3,4}, Douglas Altman¹¹, José-Antonio González¹ and Erik Cobo^{1,2}

Abstract

Background: From 2005 to 2010, we conducted 2 randomized studies on a journal (*Medicina Clínica*), where we took manuscripts received for publication and randomly assigned them to either the standard editorial process or to additional processes. Both studies were based on the use of methodological reviewers and reporting guidelines (RG). Those interventions slightly improved the items reported on the Manuscript Quality Assessment Instrument (MQAI), which assesses the quality of the research report. However, masked evaluators were able to guess the allocated group in 62% (56/90) of the papers, thus presenting a risk of detection bias. In this post-hoc study, we analyse whether those interventions that were originally designed for improving the completeness of manuscript reporting may have had an effect on the number of citations, which is the measured outcome that we used.

Methods: Masked to the intervention group, one of us used the Web of Science (WoS) to quantify the number of citations that the participating manuscripts received up December 2016. We calculated the mean citation ratio between intervention arms and then quantified the uncertainty of it by means of the Jackknife method, which avoids assumptions about the distribution shape.

Results: Our study included 191 articles (99 and 92, respectively) from the two previous studies, which all together received 1336 citations. In both studies, the groups subjected to additional processes showed higher averages, standard deviations and annual rates. The intervention effect was similar in both studies, with a combined estimate of a 43% (95% CI: 3 to 98%) increase in the number of citations.

Conclusions: We interpret that those effects are driven mainly by introducing into the editorial process a senior methodologist to find missing RG items. Those results are promising, but not definitive due to the exploratory nature of the study and some important caveats such as: the limitations of using the number of citations as a measure of scientific impact; and the fact that our study is based on a single journal. We invite journals to perform their own studies to ascertain whether or not scientific repercussion is increased by adhering to reporting guidelines and further involving statisticians in the editorial process.

Keywords: Reporting guidelines, Peer-review, Reproducibility, Transparency, Number of citations

* Correspondence: marta.vilaro@upc.edu

¹Universitat Politècnica Catalunya, Barcelona, Spain

¹²Statistical Researcher, Statistics and Operational Research, Barcelona Tech,

C/Jordi Girona, 1-3. Edifici C5, planta 2, Campus Nord, 08034 Barcelona, Spain

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

Background

The full progress of science relies on peer review, yet many have called into question the benefits of peer review [1–7]. In essence, critics assert that “studies have shown that peer reviewers were not able to appropriately detect errors, improve the completeness of reporting, or decrease the distortion of the study results” [8]. Nevertheless, the purposes for which Reporting Guidelines (RG) have been developed over the past two decades are to help authors, editors and peer reviewers check and improve the transparency of research studies while ensuring that papers are both accurate and complete [9–14]. According to the systematic review published by Bruce et al. (2016) [8], which we expand on in Section 7 of the Additional file 1, at least 23 randomized trials have studied some aspects of the peer review process, with the majority of them focusing on the quality of peer review as a surrogate outcome while only 3 [15–17] analysed the completeness of reporting as an outcome. Of these 3 trials that we previously conducted, only 2 [15, 16] found positive results regarding completeness of reporting — although only one of these reached statistical significance. Those studies were based on a partly subjective outcome, the Manuscript Quality Assessment Instrument (MQAI) [18], and there is evidence that evaluators could have successfully guessed which were in the intervention group [8]. Consequently, raters could have favoured the group receiving an additional intervention, thus raising the risk of detection bias. Therefore, we follow up on those studies here by taking advantage of the Web of Science [19] (WoS) to reassess those 2 trials by using the number of citations later received by those papers. We consider such a measured outcome to be impartial and fair, as it is naturally free from the risk of evaluation bias. The relationship between the completeness of a report and the number of citations has been previously studied, with promising though not statistically significant results having been found [20, 21]. We also previously explored this relationship with a shorter follow-up (SM, Sections 5 and 6).

Methods

We conducted two previous trials [15, 16], in which we found partially positive results from adding statistical reviewers and RGs to the peer review process. The first one was conducted in 2007 and called the “Improve Quality” (IQ) study [15], in which we randomly allocated 129 suitable manuscripts into 4 intervention groups (Fig. 1a). Unfortunately, after peer review, 16 manuscripts were rejected and 14 were lost to follow-up. Those losses introduced unpredictable (attrition) bias [22, 23] and may have affected the estimates.

The second trial was the 2011 “Enhance Transparency” (ET) study [16], in which we randomized 92 manuscripts either to both a statistical review and RGs or to neither

(Fig. 1b). In both the IQ and ET studies, the main outcome was an assessed rather than measured endpoint. As masked evaluators were able to guess the intervention arm more often than could be ascribed to chance, partially unblinded evaluators could have introduced detection bias in both studies [8].

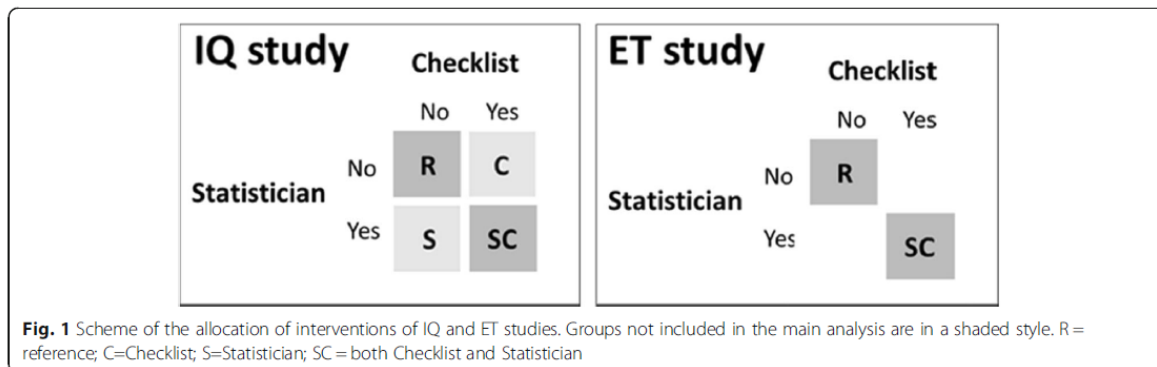
Due to these limitations, and in order to assess the long-term impact of those interventions, we adopted a new main outcome: the number of citations that each paper received on the WoS from publication up to December 31, 2016, with our hypothesis being that greater transparency and more comprehensible reporting may facilitate an increase in citations.

The IQ study divided the papers into 4 groups as a result of combining the two interventions into a 2×2 factorial design: a suggestion to the reviewers to employ an abridged checklist for the evaluation of basic biomedical research papers (C) [24]; and adding a statistician (S) from the reviewer panel list. Consequently, the 4 groups were defined as: papers which received a standard review process (reference), papers which received a review process using a local checklist (C), papers which received a standard review process and a revision from a statistician (S) and papers which received a standard review process and a revision from a statistician using a local checklist (SC). The reference intervention followed the usual journal process based on 1–3 reviewers. In order to combine those results with those of the ET study, only the 24 papers allocated to the group with both interventions (C and S) and the 27 allocated to the reference group (neither C nor S) were now included in the main analysis.

The ET study modified this design in 3 ways: first, by relying on just one senior methodological expert rather than choosing a statistical reviewer from an expert list; second, by combining both interventions, with the senior methodological reviewer proposing specific changes based on relevant international reporting guidelines; and, third, it avoided attrition by delaying the intervention until the decision had been made on whether or not to publish.

Masked to the intervention group, one of us (MV) collected from WoS the number of citations that the ET and IQ articles received. A search was made using the website’s search tab and including 3 references: (1) the publication name, “Medicina Clínica (Barcelona)”; (2) the publication year (either 2004 to 2005 or 2009 to 2010); and, (3) either the article’s title or by searching for the topic in order to consider posterior changes to the title (between the submitted and finally published version). Baseline MQAI and study group were obtained from the data of the ET and IQ studies.

We aim to estimate the ratio of the average citation-per-year between intervention arms (which we



refer to in this paper as “mean citation ratio”). As the data did not fit to the distributional assumptions of the previously masked specified Poisson model, our main analysis relies on the more robust Jackknife method, which provides wider and more conservative intervals. As sensitivity analyses, we also report alternative analyses such as the previously mentioned Poisson model (Sections 2 to 4 of SM).

Additional collected variables are described in Section 1 of SM. Section 6 of SM and the master’s thesis of the first author [25] show the results of other exploratory data analyses that were previously performed with shorter follow-up.

Analyses were performed using R software version 3.2.1.

Availability of data and materials

The dataset supporting the conclusions of this article is available at <https://www-eio.upc.edu/redir/Number Citations>, where researchers can: (1) reproduce the results of our analysis; (2) check our data at the Web of Science [19] as of December 2016; and (3) update the number of citations in order to replicate our results with a longer follow-up. The critical scientist can try to reproduce both our outcome measurements and analyses.

Results

Of the 129 randomized papers, 99 IQ articles were published between 4 February 2005 and 12 May 2006, with a mean (standard deviation (SD)) follow-up period of 11.35 (0.31) years. Those publications received a total of 927 citations (mean 9.36, SD 14.87). ET included 92 randomized papers that were published between 24 June 2009 and 3 April 2010, with a mean (SD) follow-up period of 7.29 (0.31) years. They received a total of 409 citations (mean 4.44, SD 4.08). In both studies, the group with both interventions had larger means, standard deviations and annual rates. All intervention groups

also had a slightly increased number of articles with 0 citations (Table 1 and Fig. 2).

Figure 3 shows positive intervention effects that are similar in both studies. Combining both estimates, the intervention increased the citation rate by 43% (95%CI: 3 to 98%). This effect is due mainly to the ET study, which has higher weight (85.9) in the meta-analysis due to a more precise estimate. The weight of the studies within the meta-analysis has been calculated from the inverse of the variances of mean ratio estimates, thereby obtaining 31.58 and 5.17 for ET and IQ, respectively.

All analyses show some intervention effect (Fig. 3), which is slightly larger in the ET study while there is greater uncertainty from random error in the IQ study.

In order to check the robustness of the results, we ran sensitivity analyses: a pre-specified Poisson model (which provided shorter and consequently more-precise confidence intervals); and alternative statistical models that were suitable for count data (Sections 2 to 4 in SM). All together, these provided consistent results.

Discussion

If we consider both studies together, we find that including a methodological reviewer (for example, a senior statistician) who is dedicated to looking for missing RG items increases the number of citations by 43% (95%CI: 3 to 98%), a result that — if this finding is sustained — might justify the cost and time expenditure by the journal [26, 27]. The number of papers with zero-citations was also higher in the intervention groups of both studies, which raises the possibility that greater transparency deters citations for some kinds of papers. This unexpected result warrants confirmation in future studies.

To the best of our knowledge, this is the first study showing that the completeness of reporting is potentially associated with higher citation counts as a result of a specific intervention, namely: adding to the peer review process a methodological expert who ensures that the reporting guidelines are adhered to. Although the number of citations is considered by some authors to be an

Table 1 Number of citations by study and intervention group

		N ¹	N ²	Number of citations		Annual rate Mean (SD)	Articles with 0 citations
				Mean (SD)	Median (Max)		
IQ study	Standard review process (reference)	37	27	8.4 (12.2)	4 (45)	0.7 (1.1)	1 (3.7%)
	Statistician	31	26	8.4 (13.7)	4.5 (67)	0.7 (1.2)	4 (15.4%)
	Checklist	32	22	10.3 (18.8)	4.5 (89)	0.9 (1.6)	3 (13.6%)
	Statistician + Checklist	29	24	10.7 (15.5)	6.5 (60)	0.9 (1.3)	3 (12.5%)
ET study	Standard review process (reference)	41	41	3.6 (2.5)	3 (10)	0.5 (0.3)	2 (4.9%)
	Statistician + Checklist	51	51	5.1 (4.9)	3 (19)	0.7 (0.7)	7 (13.7%)

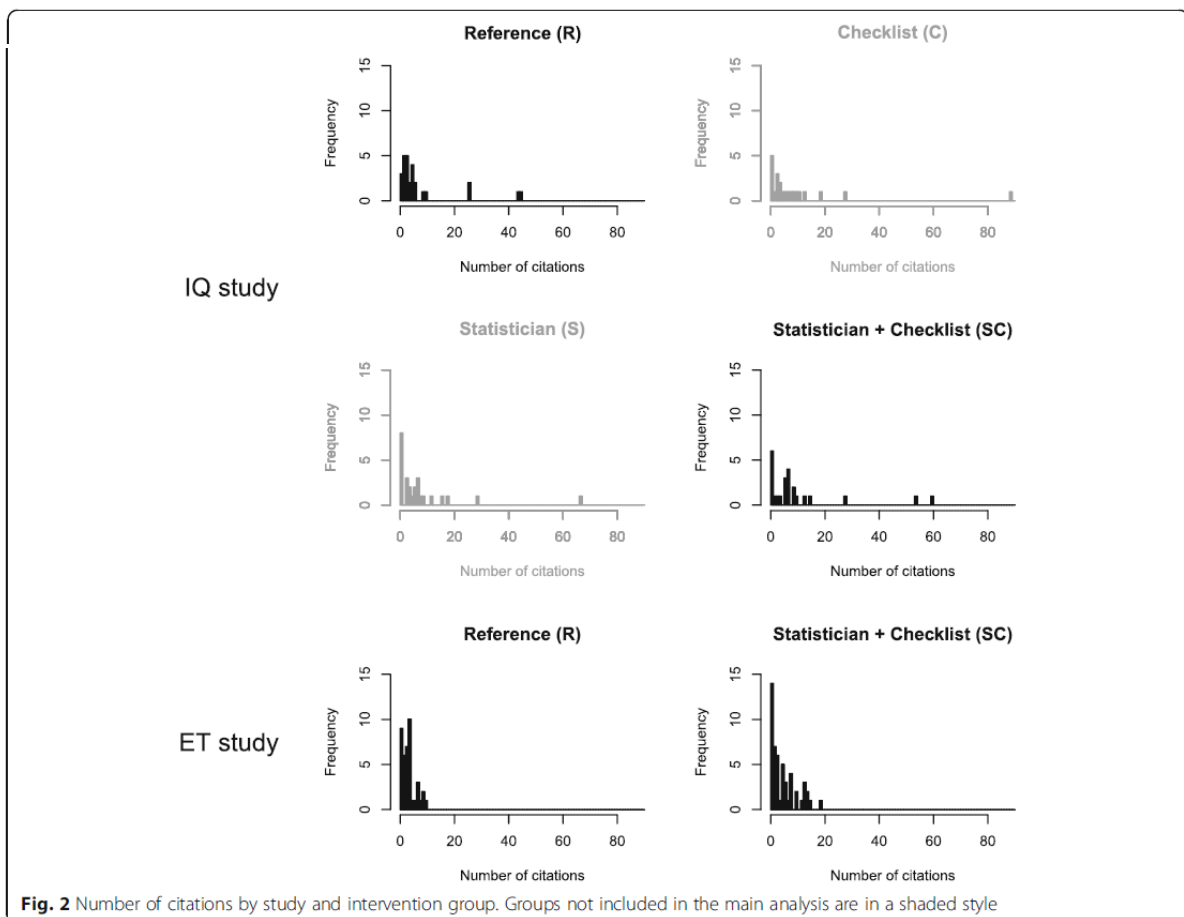
N¹ = number of randomized manuscripts; N² = number of analysed manuscripts

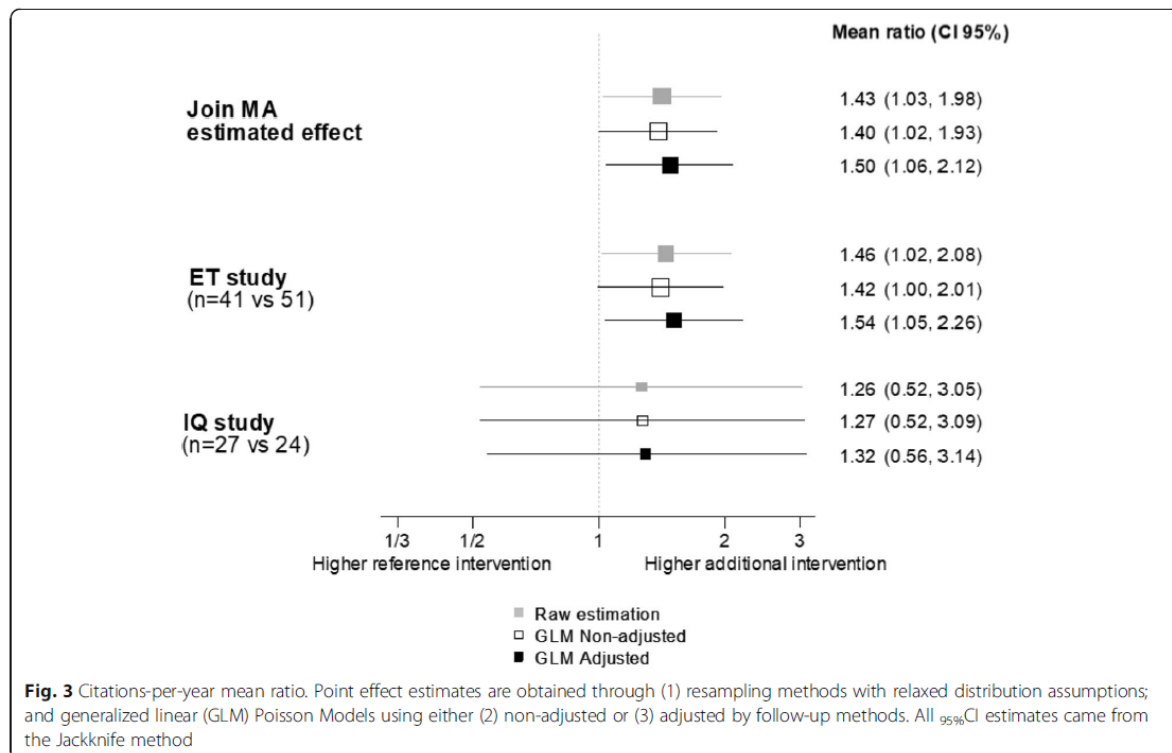
indicator of a paper's influence [28–32], some have argued that “citation counts are an indicator more of impact than of quality or importance” [28]; thus, we should not conflate the number of citations with research quality [21, 33]. Due to the high uncertainty behind the IQ study (including the risk of bias due to attrition) and the weight of the ET study when estimating the combined effects, our interpretation mainly follows the ET results in that the formal use

of RGs at the end of the editorial phase, after peer review, leads to an increase in the papers' potential scientific value. This interpretation assumes that all added citations are “positive” in the sense that they contribute to the body of knowledge.

In interpreting this effect size, we should keep in mind the uncertainty reflected by the confidence intervals.

Our next important limitation pertains to the fact that our results rely on just one journal that is not





top-quality and they therefore cannot be transported to top-tier journals where those interventions have probably already been implemented. According to the Scimago Journal Country Rank website, journals with Impact Factor ≥ 10 account for just 1% (15,259 out of 1,528,749 articles published in 2016) of biomedical scientific production; thus, our focus is not on the top-quality journals but on second-tier journals who could benefit from the intervention.

It is essential that our results be interpreted according to the exploratory nature of this extended follow-up study. First, we did not have enough advance information to know the fit between our data and the statistical models. Second, and more importantly, we had neither previous studies to sustain the hypothesis nor a sample size rationale to guarantee any desired power for testing this hypothesis. Therefore, in keeping with the American Statistical Association (ASA) statement on p -value [34], we should not interpret the results of any hypothesis test. Accordingly, we should also not be concerned about whether or not the 95% confidence intervals (CI) include the neutral value of 1, because there is no such previous hypothesis. However, as we stated prior to data collection that our objective is “to estimate the effects of those interventions on the number of citations”, selective outcome reporting is of no concern.

Conclusions

Our findings indicate that the citation counts increased by 43% (95% CI from: 3 to 98%) after including in the editorial process a methodologist who ensures the proper reporting of checklist items. As our original studies were originally designed to test those hypotheses for a different outcome, this present study was not powered to test this post-hoc analysis; therefore, our results should not be interpreted as definitive and they need to be confirmed in properly powered designs. We invite journals to perform their own studies to ascertain whether or not scientific impact is increased, first, by adhering to reporting guidelines, and second, by further involving statisticians or methodological experts in the editorial process.

Additional file

Additional file 1: Adherence to reporting guidelines increases the number of citations: the argument for including a methodologist in the editorial process and peer-review. Supplementary material. (DOCX 687 kb)

Abbreviations

ASA: American Statistician Association; C: Intervention group with suggestion to the reviewers to employ an abridged checklist for the evaluation of basic biomedical research papers; CI: Confidence Interval; ET: “Enhance Transparency” study. Cobo E, Selva-O’Callaghan A, Ribera JM, Cardellach F, Dominguez R, Vilardell M. Statistical Reviewers Improve Reporting in Biomedical Articles: A Randomized Trial. Plos One. 2007; 2 (3): e332; IQ: “Improve

Quality" study. Cobo E, Cortés J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomized trial. *BMJ*. 2011; 343: d6783; MQAI: Manuscript Quality Assessment Instrument; RG: Reporting Guidelines; S: Intervention group adding a Statistician from the reviewer panel list statistician; SD: Standard Deviation; SM: Supplementary Material; WoS: Web of Science

Acknowledgements

Not applicable.

Funding

DGA, JAG and EC belong to MiRoR (Methods in Research on Research), which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 676207. EC and JC are partially supported by the grants MTM2015-64465-C2-1-R (MINECO/FEDER, UE) from the Secretaría de Estado de Investigación, Desarrollo e Innovación del Ministerio de Economía y Competitividad (Spain) and through grant 2017 SGR 622 from the Departament d'Economia i Coneixement de la Generalitat de Catalunya (Spain). The funding bodies had no role in the study design, data collection and analysis, results interpretation, or in writing the manuscript.

Availability of data and materials

The dataset supporting the conclusions of this article is available at <https://www.eio.upcedu/redir/NumberCitations>, where researchers can: (1) reproduce the results of our analysis; (2) check our data at the Web of Science [20] as of December 2016; and (3) update the number of citations in order to replicate our results with a longer follow-up. The critical scientist can try to reproduce both our outcome measurements and analyses.

Authors' contributions

All authors contributed to and approved the paper, drafted by MV, EC, DGA, JAG, ME, JC and XB. MV collected and analysed the citation data. JAG and EC had the original idea. JC analysed the previous ET study. DGA co-designed the ET study; JMR, FC, AS, AU and MV designed and conducted both original studies and contributed to the new idea. XB validated the statistical code and results.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Universitat Politècnica Catalunya, Barcelona, Spain. ²Medicina Clínica, Elsevier-Barcelona, Barcelona, Spain. ³Universitat Autònoma de Barcelona, Barcelona, Spain. ⁴Vall D'Hebron Hospital, Barcelona, Spain. ⁵Hospital Germans Trias i Pujol, Badalona, Spain. ⁶José Carreras Leukaemia Research Institute, Catalan Institute of Oncology, Badalona, Spain. ⁷Universitat de Barcelona and Hospital Clínic, Barcelona, Spain. ⁸ISGlobal, Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain. ⁹Universitat Pompeu Fabra, Barcelona, Spain. ¹⁰Ciber Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain. ¹¹Centre for Statistics in Medicine, University of Oxford, Botnar Research Centre, Oxford, UK. ¹²Statistical Researcher, Statistics and Operational Research, Barcelona Tech, C/Jordi Girona, 1-3. Edifici C5, planta 2, Campus Nord, 08034 Barcelona, Spain.

Received: 22 October 2018 Accepted: 29 April 2019

Published online: 31 May 2019

References

- Smith R. Peer review: a flawed process at the heart of science and journals. *J R Soc Med*. 2006;99(4):178–82.
- Baxt WG, Waeckerle JF, Berlin JA, et al. Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. *Ann Emerg Med*. 1998;32(3 Pt 1):310–7.
- Shashok K. Content and communication: how can peer review provide helpful feedback about the writing. *BMC Med Res Methodol*. 2008;8:3.
- Kravitz RL, Franks P, Feldman MD, et al. Editorial peer reviewers' recommendations at a general medical journal: are they reliable and do editors care? *PLoS One*. 2010;5(4):e10072.
- Henderson M. Problems with peer review. *BMJ*. 2010;340:c1409.
- Yaffe MB. Re-reviewing peer review. *Sciolo Signal*. 2009;2(85):eg11.
- Stahel PF, Moore EE. Peer review for biomedical publications: we can improve the system. *BMC Med*. 2014;12(1):179.
- Bruce R, Chauvin A, Trinquart L, Ravaut P, Boutron I. Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Med*. 2016;14:85.
- Wang X, Chen Y, Yang N, et al. Methodology and reporting quality of reporting guidelines: systematic review. *BMC Med Res Methodol*. 2015;15:74.
- Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR network. *BMC Med*. 2010;8:24.
- McIntosh L, Juehne A, Vitale C, et al. Repeat: a framework to assess empirical reproducibility in biomedical research. *BMC Med Res Methodol*. 2017;17:143.
- Park I, Peacey MW, Munafo MR. Modelling the effects of subjective and objective decision making in scientific peer review. *Nature*. 2014;506(7486):93–6.
- Triaridis S, Kyrgidis A. Peer review and journal impact factor: the two pillars of contemporary medical publishing. *Hippokratia*. 2010;14(Suppl 1):5–12.
- Shamseer L, Stevens A, Skidmore B, et al. Does journal endorsement of reporting guidelines influence the completeness of reporting of health research? A systematic review protocol. *Syst Rev*. 2012;1:24.
- Cobo E, Selva-O'Callaghan A, Ribera JM, Cardellach F, Dominguez R, Vilardell M. Statistical reviewers improve reporting in biomedical articles: a randomized trial. *PLoS One*. 2007;2(3):e332.
- Cobo E, Cortés J, Ribera JM, et al. Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical journal: masked randomized trial. *BMJ*. 2011;343:d6783.
- Arnau C, Cobo E, Ribera JM, et al. Effect of statistical review on manuscript quality in Medicina Clínica (Barcelona): a randomized study. *Med Clin (Barc)*. 2003;121(18):690–4. [https://doi.org/10.1016/S0025-7753\(03\)74064-0](https://doi.org/10.1016/S0025-7753(03)74064-0).
- Goodman SN, Berlin J, Fletcher SW, Fletcher RH. Manuscript quality before and after peer review and editing at annals of internal medicine. *Ann Intern Med*. 1994;121(1):11–21.
- Web of Science. <https://webofknowledge.com>. Accessed 1 Jan 2017.
- Van der Pol CB, McInnes MDF, Petrich W, Tunis AS, Hanna R. Is quality and completeness of reporting of systematic reviews and meta-analyses published in high impact radiology journals associated with citation rates? *PLoS One*. 2015;10(3):e0119892. <https://doi.org/10.1371/journal.pone.0119892>.
- Nieminem P, Carpenter J, Rucker G, Shumacher M. The relationship between quality of research and citation frequency. *BMC Med Res Methodol*. 2006;6:42. <https://doi.org/10.1186/1471-2288-6-42>.
- Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. *J Clin Epidemiol*. 2007;60(7):663–9.
- Hernan M, Alonso A, Logan R, et al. Observational studies analysed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19(6):766–79.
- Bosch F, Guardiola E. Abridged checklist for the evaluation of basic biomedical research papers. *Med Clin*. 2003;121(6):228–30.
- Vilaró M. Master thesis: improving the impact of biomedical research. 2015. <https://www.hdl.handle.net/2099.1/26123>.
- Chan AW, et al. Increasing value and reducing waste: addressing inaccessible research. *The Lancet*. 2014;383(9913):257–66.
- Glaziov P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267–76.

28. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J Am Soc Inf Sci.* 1989;40:342–9.
29. Subelj L, Fiala D, Bajec M. Network-based statistical comparison of citation topology of bibliographic databases. *Nature.* 2014;4:6496.
30. Yao L, Wei T, Zeng A, Fan Y, Di Z. Ranking scientific publications: the effect of nonlinearity. *Nature.* 2014;4:6663.
31. Zhang Z, Van Poucke S. Citations for randomized controlled trials in sepsis literature: the halo effect caused by journal impact factor. *PLoS One.* 2017;12(1):e0169398.
32. Martin BR. The use of multiple indicators in the assessment of basic research. *Scientometrics.* 1996;36(3):343–62.
33. Seglen PO. The skewness of science. *J Am Soc Inf Sci.* 1992;43(9):628–38.
34. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat.* 2016;70(2):129–33.

ANNEX K. R code for the analyses

```
#####
###
### R CODE FOR MAIN COMPARISON
###
#####

rm(list=ls())

library(rmeta)
library(pscl)
windowsFonts("Times New Roman" = windowsFont("Times New Roman"))

##-- Data collected in 31/12/2018
datos <- read.csv('data_2018.csv',header=TRUE,sep=";")
datos$Follow.up <- as.numeric((as.Date("31/12/2018", "%d/%m/%Y")-
as.Date(datos$Publication.date, "%d/%m/%Y"))/365)
datos1 <- datos[datos$Study=="ET",]
datos2 <- datos[datos$Study=="IQ" & (datos$Group=="R" | datos$Group=="SC"),]

#- RAW ESTIMATION JACKKNIFE
log.mean.ratio <- function(dat) {
  d=dat; m0=mean(d$response[d$group=="R"],na.rm=T);
  m1=mean(d$response[d$group=="SC"],na.rm=T); log(m1/m0)
} # Function to find log(mean ratio)
# ET study
response <- datos1$NC/datos1$Follow.up;group <- datos1$Group
dat <- data.frame(response,group=as.factor(group)); estim <- c(); estim[1] <- log.mean.ratio(dat)
for (i in 1:dim(dat)[1]){ d <- dat[-i,]; estim[i] <- log.mean.ratio(d)
n <- dim(dat)[1]; desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
ET.jac <- round(exp(mean(estim)),2); se.ET.jac <- desv
LL.ET.jac <- round(exp(mean(estim)-qt(0.975,dim(datos[datos$Study=="ET",]) [1]-1)*desv),2)
UL.ET.jac <- round(exp(mean(estim)+qt(0.975,dim(datos[datos$Study=="ET",]) [1]-1)*desv),2)
ET_row_jac <- data.frame(x=c("Row JK ET: R vs
SC"),y=round(c(ET.jac),2),ylo=round(c(LL.ET.jac),2),yhi=round(c(UL.ET.jac),2),se=se.ET.jac)
# IQ study: Both interventions effect (R vs SC)
response <- datos2$NC/datos2$Follow.up
group <- datos2$Group; dat <- data.frame(response,group=as.factor(group))
estim <- c(); estim[1] <- log.mean.ratio(dat);
for (i in 1:dim(dat)[1]){d <- dat[-i,]; estim[i] <- log.mean.ratio(d)
n <- dim(dat)[1]; desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
IQ.jac <- round(exp(mean(estim)),2); se.IQ.jac <- desv
LL.IQ.jac <- round(exp(mean(estim)-qt(0.975,dim(datos[datos$Study=="IQ" & (datos$Group=="R" |
datos$Group=="SC"),]) [1]-1)*desv),2)
UL.IQ.jac <- round(exp(mean(estim)+qt(0.975,dim(datos[datos$Study=="IQ" & (datos$Group=="R" |
datos$Group=="SC"),]) [1]-1)*desv),2)
IQ_row_jac <- data.frame(x=c("Row JK IQ: R vs
SC"),y=round(c(IQ.jac),2),ylo=round(c(LL.IQ.jac),2),yhi=round(c(UL.IQ.jac),2),se=se.IQ.jac)
# Meta-analysis
logs <- c(log(ET_row_jac$y),log(IQ_row_jac$y)); selogs <- c(ET_row_jac$se,IQ_row_jac$se); names
<- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs); d <- meta.summaries(b$logs, b$selogs, names=b$names,
method="fix", logscale=TRUE)
estim.meta.2018 <- round(exp(d$summary),2); estim.meta.se.2018 <- round(d$se.summary,2)
LL.estim.meta.2018 <- round(summary(d)$sumnci[1],2); UL.estim.meta.2018 <-
round(summary(d)$sumnci[3],2)
MA_row_jac <- data.frame(x=c("Row JK MA: R vs
SC"),y=round(c(estim.meta.2018),2),ylo=round(c(LL.estim.meta.2018),2),
yhi=round(c(UL.estim.meta.2018),2),se=estim.meta.se.2018)
# Join estimation (no MA)
response <- datos$NC/datos$Follow.up;group <- datos$Group
dat <- data.frame(response,group=as.factor(group)); estim <- c(); estim[1] <- log.mean.ratio(dat)
for (i in 1:dim(dat)[1]){ d <- dat[-i,]; estim[i] <- log.mean.ratio(d)
n <- dim(dat)[1]; desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
Join.jac <- round(exp(mean(estim)),2); se.Join.jac <- desv
LL.Join.jac <- round(exp(mean(estim)-qt(0.975,dim(datos)[1]-1)*desv),2)
UL.Join.jac <- round(exp(mean(estim)+qt(0.975,dim(datos)[1]-1)*desv),2)
Join_row_jac <- data.frame(x=c("Row join: R vs
SC"),y=round(c(Join.jac),2),ylo=round(c(LL.Join.jac),2),
yhi=round(c(UL.Join.jac),2),se=se.Join.jac)

#- POISSON GLM JACKKNIFE
# ET: Adjusted Poisson (offset(FU) + Baseline Goodman scale)
```

```

S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5.,
data = datos1,family="poisson"));
estim <- c(); estim[1] <- S$coefficients[2,1]
windows(); par(mfrow=c(2,2)); plot(m1,main="Adjusted ET")
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(m1 <- glm(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data =
d,family="poisson"));
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos1)[1]; desv <-sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
ET_adj.pois <- exp(mean(estim)); se.Adj.ET.pois <- desv
ET_adj_pois_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv);
ET_adj_pois_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj_pois <- data.frame(x=c("Adj Pois ET: R vs SC"),y=round(c(ET_adj.pois),2),

ylo=round(c(ET_adj_pois_inf),2),yhi=round(c(ET_adj_pois_sup),2),se=se.Adj.ET.pois)
# ET: Non-Adjusted Poisson
S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data = datos1,family="poisson"));
estim <- c(); estim[1] <- S$coefficients[2,1]
windows(); par(mfrow=c(2,2)); plot(m1)
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,];
  S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data = d,family="poisson"))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos1)[1]; desv <-sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
ET_no_adj.pois <- exp(mean(estim)); se.no.Adj.ET.pois <- desv
ET_no_adj_pois_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv);
ET_no_adj_pois_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj_pois <- data.frame(x=c("No Adj Pois ET: R vs SC"),y=round(c(ET_no_adj.pois),2),
ylo=round(c(ET_no_adj_pois_inf),2),yhi=round(c(ET_no_adj_pois_sup),2),se=se.no.Adj.ET.pois)
# IQ: Adjusted Poisson (offset(FU) + Baseline Goodman scale)
S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5.,
data = datos2,family="poisson"))
estim <- c(); estim[1] <- S$coefficients[2,1];
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,];
  S <- summary(m1 <- glm(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = d,family="poisson"))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos2)[1]; desv <-sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
IQ_adj.pois <- exp(mean(estim)); se.Adj.IQ.pois <- desv
IQ_adj_pois_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv);
IQ_adj_pois_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj_pois <- data.frame(x=c("Adj Pois IQ: R vs
SC"),y=round(c(IQ_adj.pois),2),ylo=round(c(IQ_adj_pois_inf),2),
yhi=round(c(IQ_adj_pois_sup),2),se=se.Adj.IQ.pois)
# IQ: Non-Adjusted Poisson
S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data = datos2,family="poisson"));
estim <- c(); estim[1] <- S$coefficients[2,1]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,];
  S <- summary(m1 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data = d,family="poisson"));
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos2)[1]; desv <-sqrt(((n-1)/n)*sum((estim-mean(estim))^2));
IQ_no_adj.pois <- exp(mean(estim)); se.no.Adj.IQ.pois <- desv
IQ_no_adj_pois_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv);
IQ_no_adj_pois_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj_pois <- data.frame(x=c("No Adj Pois IQ: R vs SC"),y=round(c(IQ_no_adj.pois),2),

ylo=round(c(IQ_no_adj_pois_inf),2),yhi=round(c(IQ_no_adj_pois_sup),2),se=se.no.Adj.IQ.pois)
# Meta-analysis
# Unadjusted
logs <- c(log(ET_no_adj_pois$y),log(IQ_no_adj_pois$y)); selogs <-
c(ET_no_adj_pois$se,IQ_no_adj_pois$se)
names <- c("ET study", "IQ study"); b <- data.frame(names,logs,selogs); d <-
meta.summaries(b$logs, b$selogs,
names=b$names,method="fix", logscale=TRUE)
estim.no.Adj.meta <- round(exp(d$summary),2); LL.estim.no.Adj.meta <-
round(summary(d)$sumnci[1],2);
UL.estim.no.Adj.meta <- round(summary(d)$sumnci[3],2)
se.estim.meta.no.Adj <- d$se.summary

```

```

MA_no_adj_pois <- data.frame(x=c("No Adj Pois MA: R vs
SC"),y=round(c(estim.no.Adj.meta),2),ylo=round(c(LL.estim.no.Adj.meta),2),yhi=round(c(UL.estim.no
.Adj.meta),2),se=se.estim.meta.no.Adj)
# Adjusted
logs <- c(log(ET_adj_pois$y),log(IQ_adj_pois$y)); selogs <- c(ET_adj_pois$se,IQ_adj_pois$se)
d <- meta.summaries(logs, selogs, names=names,method="fix", logscale=TRUE); estim.Adj.meta <-
round(exp(d$summary),2)
LL.estim.Adj.meta <- round(summary(d)$summci[1],2);
UL.estim.Adj.meta <- round(summary(d)$summci[3],2); se.estim.meta.Adj <- d$se.summary
MA_adj_pois <- data.frame(x=c("Adj Pois MA: R vs
SC"),y=round(c(estim.Adj.meta),2),ylo=round(c(LL.estim.Adj.meta),2),yhi=round(c(UL.estim.Adj.meta
),2),se=se.estim.meta.Adj)

# Elements for the graphic
MA_row_jac
MA_no_adj_pois
MA_adj_pois
ET_row_jac
ET_no_adj_pois
ET_adj_pois
IQ_row_jac
IQ_no_adj_pois
IQ_adj_pois

# GRAHIC - Forest plot (se JK)
library(export)
win.graph(8,6); par(oma=c(0,6,0,9)); par(family = "Times New Roman")
poin <- c(22,22); lin <- c(1,1); wid <- c(1,2); pos <- c(4:1); etiq <- c("(n=41 vs 51)","(n=27 vs
24)")
plot(NA, xlim = c(0.33,3),
ylim=c(0,pos[1]),xaxt="n",yaxt="n",ylab="",xlab="",axes=F,log="x",main="",font.main=4)
abline(v=1,lty=3,col="darkgrey"); axis(1,c(1/3.5,0.33,0.5,1,2,3,3.5),labels=F);
mtext(c("1/3","1/2","1","2","3"),las=1,side=1,at=c(0.33,0.5,1,2,3),line=0.5)
# MA
segments(MA_row_jac$ylo,pos[2]+0.8,MA_row_jac$yhi,pos[2]+0.8,lwd=wid[1],col="darkgrey",lty=1);
points(MA_row_jac$y,pos[2]+0.8,pch=poin[2],cex=0.8/sqrt(MA_row_jac$se),bg="darkgrey",col="darkgre
y")
segments(MA_no_adj_pois$ylo,pos[2]+0.5,MA_no_adj_pois$yhi,pos[2]+0.5,lwd=wid[1],lty=1);
points(MA_no_adj_pois$y,pos[2]+0.5,pch=poin[1],cex=0.8/sqrt(MA_no_adj_pois$se))
segments(MA_adj_pois$ylo,pos[2]+0.2,MA_adj_pois$yhi,pos[2]+0.2,lwd=wid[1],lty=1);
points(MA_adj_pois$y,pos[2]+0.2,pch=poin[2],cex=0.8/sqrt(MA_adj_pois$se),bg="black",col="black")
# ET
segments(ET_row_jac$ylo,pos[3]+0.2,ET_row_jac$yhi,pos[3]+0.2,lwd=wid[1],lty=1,col="darkgrey");
points(ET_row_jac$y,pos[3]+0.2,pch=poin[1],cex=0.8/sqrt(ET_row_jac$se),bg="darkgrey",col="darkgre
y")
segments(ET_no_adj_pois$ylo,pos[3]-0.1,ET_no_adj_pois$yhi,pos[3]-0.1,lty=1);
points(ET_no_adj_pois$y,pos[3]-0.1,pch=poin[2],cex=0.8/sqrt(ET_no_adj_pois$se))
segments(ET_adj_pois$ylo,pos[3]-0.4,ET_adj_pois$yhi ,pos[3]-0.4,lwd=wid[1],lty=1);
points(ET_adj_pois$y,pos[3]-0.4,pch=poin[2],cex=0.8/sqrt(ET_adj_pois$se),bg="black",col="black")
# IQ
segments(IQ_row_jac$ylo,pos[4],IQ_row_jac$yhi,pos[4],lwd=wid[1],lty=1,col="darkgrey");
points(IQ_row_jac$y,pos[4],pch=poin[1],cex=0.8/sqrt(IQ_row_jac$se),bg="darkgrey",col="darkgrey")
segments(IQ_no_adj_pois$ylo,pos[4]-0.3,IQ_no_adj_pois$yhi,pos[4]-0.3,lwd=wid[1],lty=1);
points(IQ_no_adj_pois$y,pos[4]-0.3,pch=poin[2],cex=0.8/sqrt(IQ_no_adj_pois$se))
segments(IQ_adj_pois$ylo ,pos[4]-0.6,IQ_adj_pois$yhi ,pos[4]-0.6,lwd=wid[1],lty=1);
points(IQ_adj_pois$y,pos[4]-0.6,pch=poin[2],cex=0.8/sqrt(IQ_adj_pois$se),bg="black",col="black")
# Labels
mtext("Joint MA",las=2,side=2,at=pos[2]+0.7,line=4,font=2,cex=1.2);
mtext("estimated effect",las=2,side=2,at=pos[2]+0.5,line=1.3,font=2,cex=1.2)
mtext("ET study",las=2,side=2,at=pos[3],line=3.8,font=2,cex = 1);
mtext(etiq[1],las=2,side=2,at=pos[3]-0.2,line=2.9,cex = 1)
mtext("IQ study",las=2,side=2,at=pos[4]-0.3,line=4.1,font=2,cex = 1);
mtext(etiq[2],las=2,side=2,at=pos[4]-0.5,line=3.1,cex = 1)
mtext("Mean ratio (95% CI)",las=2,side=4,at=pos[1]+0.25,line=0.5,font=2)
mtext(c("1.40","1.44","1.54","1.51","1.16","1.18","1.18"),side=4,las=2,at=c(3.8,2.2,1.9,1.6,1,0.7
,0.4),line=1,cex=c(1,rep(0.9,6)))
mtext(c("1.47","1.45"),side=4,las=2,at=c(3.5,3.2),line=1)
mtext(c("(1.01, 1.94)","(1.01, 2.07)","(0.48,
2.80)"),side=4,las=2,at=c(3.8,2.2,1),line=3,cex=c(1,rep(0.9,2)))
mtext(c("(1.04, 2.09)","(1.02, 2.05)","(1.04, 2.27)","(1.03, 2.23)","(0.49,
2.83)","(0.49,2.84)"),
side=4,las=2,at=c(3.5,3.2,1.9,1.6,0.7,0.4),line=3,cex=c(1,1,rep(0.9,4)))
mtext("Higher additional intervention",side=1,at=2.5,line=1.5); mtext("Higher reference
intervention",side=1,at=1/2.5,line=1.5)
par(xpd=NA)
legend(0.23,4.9,bty="n",horiz = T,c("Raw estimate","Poisson Non-adjusted","Poisson
Adjusted"),pch=c(15,0,15),col=c("grey","black","black"))

```

```
#####
##
## SENSITIVITY ANALYSIS (POISSON WALD standard deviation estimation)
##
#####
rm(list=ls())

library(rmeta)
library(pscl)

##-- Udated data in 31 December 2018
datos <- read.csv('data_2018.csv',header=TRUE,sep=";")
summary(datos)

# Follow-up
datos$Publication.date <- as.Date(datos$Publication.date,format="%d/%m/%Y")
datos$Follow.up <- as.numeric(as.Date("2018-12-31")-datos$Publication.date)/365
datos1 <- datos[datos$Study=="ET",]
datos2 <- datos[datos$Study=="IQ" & (datos$Group=="R" | datos$Group=="SC"),]

##-- POISSON MODEL (Wald)
##-- ET: Adjusted Poisson (offset(FU) + Baseline Goodman scale)
S <- summary(m1.1 <- glm(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data =
datos1,family="poisson"))
desv <- S$coefficients[2,2]; ET_OR_adj.wald <- exp(S$coefficients[2,1])
ET_OR_adj_inf.wald <- exp(S$coefficients[2,1]-qt(0.975,dim(datos1)[1]-1)*desv)
ET_OR_adj_sup.wald <- exp(S$coefficients[2,1]+qt(0.975,dim(datos1)[1]-1)*desv)
se.Adj.ET.pois.wald <- desv
##-- IQ: Adjusted Poisson (offset(FU) + Baseline Goodman scale)
S <- summary(m1.2 <- glm(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data =
datos2,family="poisson"))
desv <- S$coefficients[2,2]; IQ_OR_adj.wald <- exp(S$coefficients[2,1])
IQ_OR_adj_inf.wald <- exp(S$coefficients[2,1]-qt(0.975,dim(datos1)[1]-1)*desv)
IQ_OR_adj_sup.wald <- exp(S$coefficients[2,1]+qt(0.975,dim(datos1)[1]-1)*desv)
se.Adj.IQ.pois.wald <- desv
##-- Meta-analysis
logs <- c(log(ET_OR_adj.wald),log(IQ_OR_adj.wald))
selogs <- c(se.Adj.ET.pois.wald,se.Adj.IQ.pois.wald)
names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.Adj.meta.wald <- round(exp(d$summary),2)
LL.estim.Adj.meta.wald <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.wald <- round(exp(d$summary+1.96*d$se.summary),2)
se.estim.meta.Adj.wald <- d$se.summary

##-- ET: Non-Adjusted Poisson (offset(FU))
S <- summary(m1.1 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data =
datos1,family="poisson"))
desv <- S$coefficients[2,2]; ET_OR_no_adj.wald <- exp(S$coefficients[2,1])
ET_OR_no_adj_inf.wald <- exp(S$coefficients[2,1]-qt(0.975,dim(datos1)[1]-1)*desv)
ET_OR_no_adj_sup.wald <- exp(S$coefficients[2,1]+qt(0.975,dim(datos1)[1]-1)*desv)
se.no.Adj.ET.pois.wald <- desv
##-- IQ: Non-Adjusted Poisson (offset(FU))
S <- summary(m1.2 <- glm(NC ~ as.factor(Group)+offset(Follow.up), data =
datos2,family="poisson"))
desv <- S$coefficients[2,2]; IQ_OR_no_adj.wald <- exp(S$coefficients[2,1])
IQ_OR_no_adj_inf.wald <- exp(S$coefficients[2,1]-qt(0.975,dim(datos1)[1]-1)*desv)
IQ_OR_no_adj_sup.wald <- exp(S$coefficients[2,1]+qt(0.975,dim(datos1)[1]-1)*desv)
se.no.Adj.IQ.pois.wald <- desv
##-- Meta-analysis
logs <- c(log(ET_OR_no_adj.wald),log(IQ_OR_no_adj.wald))
selogs <- c(se.no.Adj.ET.pois.wald,se.no.Adj.IQ.pois.wald)
names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.no.Adj.meta.wald <- round(exp(d$summary),2)
LL.estim.no.Adj.meta.wald <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no.Adj.meta.wald <- round(exp(d$summary+1.96*d$se.summary),2)
se.estim.meta.no.Adj.wald <- d$se.summary

# GRAHIC - Forest plot (Wald se)
library(extrafont)
win.graph(8,6); par(oma=c(0,6,0,9)); par(family = "Times New Roman")
```

```

poin <- c(22,22); lin <- c(1,1); wid <- c(1,2); pos <- c(3:1); etiq <- c("(n=41 vs 51)", "(n=27 vs
24)")
plot(NA, xlim = c(0.33,3),
ylim=c(0, pos[1]), xaxt="n", yaxt="n", ylab="", xlab="", axes=F, log="x", main="", font.main=4)
abline(v=1, lty=3, col="darkgrey"); axis(1, c(1/3.5, 0.33, 0.5, 1, 2, 3, 3.5), labels=F);
mtext(c("1/3", "1/2", "1", "2", "3"), las=1, side=1, at=c(0.33, 0.5, 1, 2, 3), line=0.5)
# MA
segments(LL.estim.no.Adj.meta.wald, pos[2]+0.5, UL.estim.no.Adj.meta.wald, pos[2]+0.5, lwd=wid[1], lty
=1); points(estim.no.Adj.meta.wald, pos[2]+0.5, pch=poin[1], cex=0.8/sqrt(se.estim.meta.no.Adj.wald
<- d$se.summary))
segments(LL.estim.Adj.meta.wald, pos[2]+0.2, UL.estim.Adj.meta.wald, pos[2]+0.2, lwd=wid[1], lty=1);
points(estim.Adj.meta.wald, pos[2]+0.2, pch=poin[2], cex=0.8/sqrt(se.estim.meta.Adj.wald <-
d$se.summary), bg="black", col="black")
# ET
segments(ET_OR_no_adj_inf.wald , pos[3]+0.4, ET_OR_no_adj_sup.wald, pos[3]+0.4, lwd=wid[1], lty=1);
points(ET_OR_no_adj.wald , pos[3]+0.4, pch=poin[1], cex=0.8/sqrt(se.no.Adj.ET.pois.wald
), bg="black", col="black")
segments(ET_OR_adj_inf.wald, pos[3]+0.1, ET_OR_adj_sup.wald, pos[3]+0.1, lwd=wid[1], lty=1);
points(ET_OR_adj.wald , pos[3]+0.1, pch=poin[2], cex=0.8/sqrt(se.Adj.ET.pois.wald ))
# IQ
segments(IQ_OR_no_adj_inf.wald, pos[3]-0.4, IQ_OR_no_adj_sup.wald , pos[3]-0.4, lwd=wid[1], lty=1);
points(IQ_OR_no_adj.wald , pos[3]-0.4, pch=poin[2], cex=0.8/sqrt(se.no.Adj.IQ.pois.wald ))
segments(IQ_OR_adj_inf.wald, pos[3]-0.7, IQ_OR_adj_sup.wald , pos[3]-0.7, lwd=wid[1], lty=1);
points(IQ_OR_adj.wald , pos[3]-0.7, pch=poin[1], cex=0.8/sqrt(se.Adj.IQ.pois.wald
), bg="black", col="black")
# Labels
mtext("Joint MA", las=2, side=2, at=pos[2]+0.5, line=4.1, font=2, cex=1.2); mtext("estimated
effect", las=2, side=2, at=pos[2]+0.3, line=1.5, font=2, cex=1.2)
mtext("ET study", las=2, side=2, at=pos[3]+0.4, line=3.8, font=2, cex = 1);
mtext(etiq[1], las=2, side=2, at=pos[3]+0.25, line=2.9, cex = 1)
mtext("IQ study", las=2, side=2, at=pos[3]-0.4, line=4.1, font=2, cex = 1);
mtext(etiq[2], las=2, side=2, at=pos[3]-0.55, line=3.1, cex = 1)
mtext("Mean ratio (95% CI)", las=2, side=4, at=pos[1]+0.25, line=0.5, font=2)
mtext(c("1.33 (1.17, 1.51)", "1.32 (1.17,
1.50)"), side=4, las=2, at=c(pos[2]+0.5, pos[2]+0.2), line=1, cex=1)
mtext(c("1.54 (1.27, 1.85)", "1.51 (1.25,
1.83)"), side=4, las=2, at=c(pos[3]+0.4, pos[3]+0.1), line=1, cex=0.9)
mtext(c("1.18 (1.00, 1.40)", "1.18 (1.00, 1.40)"), side=4, las=2, at=c(pos[3]-0.4, pos[3]-
0.7), line=1, cex=0.9)
mtext("Higher additional intervention", side=1, at=2.5, line=1.5); mtext("Higher reference
intervention", side=1, at=1/2.5, line=1.5)
par(xpd=NA)
legend(0.37, 3.7, bty="n", horiz = T, c("Poisson Non-adjusted", "Poisson
Adjusted"), pch=c(0, 15), col=c("black", "black"))

```

```
#####
##
## SENSITIVITY ANALYSIS (OTHER MODELS)
##
#####
rm(list=ls())
library(rmeta)
library(pscl)

##-- Udated data since 31 December 2018
datos <- read.csv('data_2018.csv',header=TRUE,sep=";")
datos$Follow.up <- as.numeric((as.Date("31/12/2018", "%d/%m/%Y")-
as.Date(datos$Publication.date, "%d/%m/%Y"))/365)
datos1 <- datos[datos$Study=="ET",]
datos2 <- datos[datos$Study=="IQ" & (datos$Group=="R" | datos$Group=="SC"),]

###---- NEGATIVE-BINOMIAL MODEL
#- ET (Adjusted)
library(MASS)
S <- summary(m2.1 <- glm.nb(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = datos1))
estim <- c(); estim[1] <- S$coefficients[2,1]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(glm.nb(NC ~ as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5.,
data = d))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos1)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2)); se.Adj.ET.nb <- desv
ET_adj_nb <- exp(mean(estim))
ET_adj_nb_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj_nb_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj_nb <- data.frame(x=c("Adj NB ET: R vs
SC"),y=round(c(ET_adj_nb),2),ylo=round(c(ET_adj_nb_inf),2),yhi=round(c(ET_adj_nb_sup),2),se=se.Ad
j.ET.nb)
#- IQ (Adjusted)
library(MASS)
S <- summary(m2.2 <- glm.nb(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = datos2))
estim <- c(); estim[1] <- S$coefficients[2,1]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(glm.nb(NC ~ as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5.,
data = d))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2)); se.Adj.IQ.nb <- desv
IQ_adj_nb <- exp(mean(estim))
IQ_adj_nb_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj_nb_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj_nb <- data.frame(x=c("Adj NB IQ: R vs
SC"),y=round(c(IQ_adj_nb),2),ylo=round(c(IQ_adj_nb_inf),2),yhi=round(c(IQ_adj_nb_sup),2),se=se.Ad
j.IQ.nb)
#- MA (Adjusted)
logs <- c(log(ET_adj_nb),log(IQ_adj_nb)); selogs <- c(se.Adj.ET.nb,se.Adj.IQ.nb)
names <- c("ET study", "IQ study"); b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names, method="fix", logscale=TRUE)
estim.Adj.meta.2 <- round(exp(d$summary),2); se.estim.meta.Adj <- d$se.summary
LL.estim.Adj.meta.2 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.2 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_adj_nb <- data.frame(x=c("Adj NB MA: R vs
SC"),y=round(c(estim.Adj.meta.2),2),ylo=round(c(LL.estim.Adj.meta.2),2),yhi=round(c(UL.estim.Adj.
meta.2),2),se=se.estim.meta.Adj)

#- ET (Non-Adjusted)
library(MASS)
S <- summary(m2.1 <- glm.nb(NC ~ as.factor(Group)+offset(Follow.up), data = datos1))
estim <- c(); estim[1] <- S$coefficients[2,1]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(glm.nb(NC ~ as.factor(Group)+offset(Follow.up), data = d))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos1)[1]
```

```

desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2)); se.no_Adj.ET.nb <- desv
ET_no_adj.nb <- exp(mean(estim))
ET_no_adj_nb_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj_nb_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj_nb <- data.frame(x=c("Non Adj NB ET: R vs
SC"),y=round(c(ET_no_adj.nb),2),ylo=round(c(ET_no_adj_nb_inf),2),yhi=round(c(ET_no_adj_nb_sup),2)
,se=se.no_Adj.ET.nb)
#- IQ (Non-Adjusted)
library(MASS)
S <- summary(m2.2 <- glm.nb(NC ~ as.factor(Group)+offset(Follow.up), data = datos2))
estim <- c(); estim[1] <- S$coefficients[2,1]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(glm.nb(NC ~ as.factor(Group)+offset(Follow.up), data = d))
  estim[i] <- S$coefficients[2,1]
}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2)); se.no_Adj.IQ.nb <- desv
IQ_no_adj.nb <- exp(mean(estim))
IQ_no_adj_nb_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj_nb_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj_nb <- data.frame(x=c("Non Adj NB IQ: R vs
SC"),y=round(c(IQ_no_adj.nb),2),ylo=round(c(IQ_no_adj_nb_inf),2),yhi=round(c(IQ_no_adj_nb_sup),2)
,se=se.no_Adj.IQ.nb)
#- MA (Non Adjusted)
logs <- c(log(ET_no_adj.nb),log(IQ_no_adj.nb)); selogs <- c(se.no_Adj.ET.nb,se.no_Adj.IQ.nb)
names <- c("ET study", "IQ study"); b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$log, b$selogs, names=b$names, method="fix", logscale=TRUE)
estim.no_Adj.meta.2 <- round(exp(d$summary),2); se.estim.meta.no_Adj <- d$se.summary
LL.estim.no_Adj.meta.2 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no_Adj.meta.2 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_no_adj_nb <- data.frame(x=c("Non Adj NB MA: R vs
SC"),y=round(c(estim.no_Adj.meta.2),2),ylo=round(c(LL.estim.no_Adj.meta.2),2),yhi=round(c(UL.esti
m.no_Adj.meta.2),2),se=se.estim.meta.no_Adj)

###----- ZERO-INFLATED POISSON MODEL
library(pscl)
#-- ET: Adjusted ZIP (offset(FU) + Baseline Goodman scale)
S <- summary(m4.1 <- zeroinfl(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = datos1))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up)
+Baseline.Goodman.Scale..likert.5., data = d))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos1)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-81]-mean(estim.zero[-81]))^2))
ET_adj.mean.zip <- exp(mean(estim)); se.Adj.ET.mean.zip <- desv
ET_adj.mean.zip_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj.mean.zip_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj.mean.zip <- data.frame(x=c("Adj mean ZIP ET: R vs
SC"),y=round(c(ET_adj.mean.zip),2),ylo=round(c(ET_adj.mean.zip_inf),2),yhi=round(c(ET_adj.mean.zi
p_sup),2),se=se.Adj.ET.mean.zip)
ET_adj.zero.zip <- exp(mean(estim.zero))
ET_adj_zero.zip_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos1)[1]-1)*desv.zero)
ET_adj_zero.zip_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos1)[1]-1)*desv.zero)
se.Adj.ET.zero.zip <- desv.zero
ET_adj_zero.zip <- data.frame(x=c("Adj zero ZIP ET: R vs
SC"),y=round(c(ET_adj.zero.zip),2),ylo=round(c(ET_adj_zero.zip_inf),2),yhi=round(c(ET_adj_zero_zi
p_sup),2),se=se.Adj.ET.zero.zip)
#-- IQ: Adjusted ZIP (offset(FU) + Baseline Goodman scale)
S <- summary(m4.1 <- zeroinfl(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = datos2))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(zeroinfl(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = d))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}

```



```

}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-81]-mean(estim.zero[-81]))^2))
IQ_adj.mean.zip <- exp(mean(estim)); se.Adj.IQ.mean.zip <- desv
IQ_adj.mean.zip_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj.mean.zip_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj.mean.zip <- data.frame(x=c("Adj mean ZIP IQ: R vs
SC"),y=round(c(IQ_adj.mean.zip),2),ylo=round(c(IQ_adj.mean.zip_inf),2),yhi=round(c(IQ_adj.mean.zip_sup),2),se=se.Adj.IQ.mean.zip)
IQ_adj.zero.zip <- exp(mean(estim.zero))
IQ_adj.zero.zip_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos2)[1]-1)*desv.zero)
IQ_adj.zero.zip_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos2)[1]-1)*desv.zero)
se.Adj.IQ.zero.zip <- desv.zero
IQ_adj.zero.zip <- data.frame(x=c("Adj zero ZIP IQ: R vs
SC"),y=round(c(IQ_adj.zero.zip),2),ylo=round(c(IQ_adj.zero.zip_inf),2),yhi=round(c(IQ_adj.zero.zip_sup),2),se=se.Adj.IQ.zero.zip)
#- MA
logs <- c(log(ET_adj.mean.zip),log(IQ_adj.mean.zip)); selogs <-
c(se.Adj.ET.mean.zip,se.Adj.IQ.mean.zip); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.Adj.meta.4 <- round(exp(d$summary),2); se.estim.meta.Adj <- d$se.summary
LL.estim.Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_adj.mean.zip <- data.frame(x=c("Adj mean ZIP MA: R vs
SC"),y=round(c(estim.Adj.meta.4),2),ylo=round(c(LL.estim.Adj.meta.4),2),yhi=round(c(UL.estim.Adj.meta.4),2),se=se.estim.meta.Adj)
logs <- c(log(ET_adj.zero.zip),log(IQ_adj.zero.zip)); selogs <-
c(se.Adj.ET.zero.zip,se.Adj.IQ.zero.zip); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.Adj.meta.4 <- round(exp(d$summary),2); se.estim.meta.Adj <- d$se.summary
LL.estim.Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_adj.zero.zip <- data.frame(x=c("Adj zero ZIP MA: R vs
SC"),y=round(c(estim.Adj.meta.4),2),ylo=round(c(LL.estim.Adj.meta.4),2),yhi=round(c(UL.estim.Adj.meta.4),2),se=se.estim.meta.Adj)

##-- ET: Non-Adjusted ZIP (offset(FU))
S <- summary(m4.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = datos1))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = d))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos1)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-81]-mean(estim.zero[-81]))^2))
ET_no_adj.mean.zip <- exp(mean(estim)); se.no_Adj.ET.mean.zip <- desv
ET_no_adj.mean.zip_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj.mean.zip_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj.mean.zip <- data.frame(x=c("Non Adj mean ZIP ET: R vs
SC"),y=round(c(ET_no_adj.mean.zip),2),
ylo=round(c(ET_no_adj.mean.zip_inf),2),yhi=round(c(ET_no_adj.mean.zip_sup),2),se=se.no_Adj.ET.mean.zip)
ET_no_adj.zero.zip <- exp(mean(estim.zero))
ET_no_adj.zero.zip_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos1)[1]-1)*desv.zero)
ET_no_adj.zero.zip_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos1)[1]-1)*desv.zero)
se.no_Adj.ET.zero.zip <- desv.zero
ET_no_adj.zero.zip <- data.frame(x=c("Non Adj zero ZIP ET: R vs
SC"),y=round(c(ET_no_adj.zero.zip),2),ylo=round(c(ET_no_adj.zero.zip_inf),2),
yhi=round(c(ET_no_adj.zero.zip_sup),2),se=se.no_Adj.ET.zero.zip)
##-- IQ: Non-Adjusted ZIP (offset(FU))
S <- summary(m4.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = datos2))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = d))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}

```

```

}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-81]-mean(estim.zero[-81]))^2))
IQ_no_adj.mean.zip <- exp(mean(estim)); se.no_Adj.IQ.mean.zip <- desv
IQ_no_adj.mean.zip_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj.mean.zip_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj.mean.zip <- data.frame(x=c("Non Adj mean ZIP IQ: R vs
SC"),y=round(c(IQ_no_adj.mean.zip),2),ylo=round(c(IQ_no_adj.mean.zip_inf),2),yhi=round(c(IQ_no_ad
j.mean.zip_sup),2),se=se.no_Adj.IQ.mean.zip)
IQ_no_adj.zero.zip <- exp(mean(estim.zero))
IQ_no_adj.zero.zip_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos2)[1]-1)*desv.zero)
IQ_no_adj.zero.zip_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos2)[1]-1)*desv.zero)
se.no_Adj.IQ.zero.zip<- desv.zero
IQ_no_adj.zero.zip <- data.frame(x=c("Non Adj zero ZIP IQ: R vs
SC"),y=round(c(IQ_no_adj.zero.zip),2),ylo=round(c(IQ_no_adj.zero.zip_inf),2),yhi=round(c(IQ_no_ad
j.zero.zip_sup),2),se=se.no_Adj.IQ.zero.zip)
#- MA
logs <- c(log(ET_no_adj.mean.zip),log(IQ_no_adj.mean.zip)); selogs <-
c(se.no_Adj.ET.mean.zip,se.no_Adj.IQ.mean.zip); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.no_Adj.meta.4 <- round(exp(d$summary),2); se.no_estim.meta.no_Adj <- d$se.summary
LL.estim.no_Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no_Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_no_adj.mean.zip <- data.frame(x=c("Non-Adj mean ZIP MA: R vs
SC"),y=round(c(estim.no_Adj.meta.4),2),
ylo=round(c(LL.estim.no_Adj.meta.4),2),yhi=round(c(UL.estim.no_Adj.meta.4),2),se=se.estim.meta.no
_Adj)
logs <- c(log(ET_no_adj.zero.zip),log(IQ_no_adj.zero.zip)); selogs <-
c(se.no_Adj.ET.zero.zip,se.no_Adj.IQ.zero.zip); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.no_Adj.meta.4 <- round(exp(d$summary),2); se.estim.meta.no_Adj <- d$se.summary
LL.estim.no_Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no_Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_no_adj.zero.zip <- data.frame(x=c("Non-Adj zero ZIP MA: R vs
SC"),y=round(c(estim.no_Adj.meta.4),2),ylo=round(c(LL.estim.no_Adj.meta.4),2),
yhi=round(c(UL.estim.no_Adj.meta.4),2),se=se.estim.meta.no_Adj)

####---- ZERO-INFLATED NEGATIVE BINOMIAL MODEL
library(pscl)
##-- ET: Adjusted ZIP (offset(FU) + Baseline Goodman scale)
S <- summary(m5.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up)
+Baseline.Goodman.Scale..likert.5., data = datos1,dist="negbin"))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up)
+Baseline.Goodman.Scale..likert.5., data = d,dist="negbin"))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos1)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-61]-mean(estim.zero[-61]))^2))
ET_adj.mean.zinb <- exp(mean(estim))
ET_adj.mean.zinb_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_adj.mean.zinb_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
se.Adj.ET.mean.zinb<- desv
ET_adj.mean.zinb <- data.frame(x=c("Adj mean ZINB ET: R vs
SC"),y=round(c(ET_adj.mean.zinb),2),ylo=round(c(ET_adj.mean.zinb_inf),2),yhi=round(c(ET_adj.mean.
zinb_sup),2),se=se.Adj.ET.mean.zinb)
ET_adj.zero.zinb <- exp(mean(estim.zero))
ET_adj.zero.zinb_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos1)[1]-1)*desv.zero)
ET_adj.zero.zinb_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos1)[1]-1)*desv.zero)
se.Adj.ET.zinb.zero<- desv.zero
ET_adj.zero.zinb <- data.frame(x=c("Adj zero ZINB ET: R vs
SC"),y=round(c(ET_adj.zero.zinb),2),ylo=round(c(ET_adj.zero.zinb_inf),2),
yhi=round(c(ET_adj.zero.zinb_sup),2),se=se.Adj.ET.zinb.zero)
##-- IQ: Adjusted ZIP (offset(FU) + Baseline Goodman scale)
S <- summary(m5.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up)
+Baseline.Goodman.Scale..likert.5., data = datos2,dist="negbin"))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]

```

```

desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(zeroinfl(NC ~
as.factor(Group)+offset(Follow.up)+Baseline.Goodman.Scale..likert.5., data = d,dist="negbin"))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-61]-mean(estim.zero[-61]))^2))
IQ_adj.mean.zinb <- exp(mean(estim))
IQ_adj.mean.zinb_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_adj.mean.zinb_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
se.Adj.IQ.mean.zinb<- desv
IQ_adj_mean.zinb <- data.frame(x=c("Adj mean ZINB IQ: R vs
SC"),y=round(c(IQ_adj.mean.zinb),2),ylo=round(c(IQ_adj.mean.zinb_inf),2),
yhi=round(c(IQ_adj.mean.zinb_sup),2),se=se.Adj.IQ.mean.zinb)
IQ_adj.zero.zinb <- exp(mean(estim.zero))
IQ_adj.zero.zinb_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos2)[1]-1)*desv.zero)
IQ_adj.zero.zinb_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos2)[1]-1)*desv.zero)
se.Adj.IQ.zinb.zero<- desv.zero
IQ_adj_zero.zinb <- data.frame(x=c("Adj zero ZINB IQ: R vs
SC"),y=round(c(IQ_adj.zero.zinb),2),ylo=round(c(IQ_adj.zero.zinb_inf),2),
yhi=round(c(IQ_adj.zero.zinb_sup),2),se=se.Adj.IQ.zinb.zero)
#- MA
logs <- c(log(ET_adj.mean.zinb),log(IQ_adj.mean.zinb)); selogs <-
c(se.Adj.ET.mean.zinb,se.Adj.IQ.mean.zinb); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.Adj.meta.4 <- round(exp(d$summary),2); se.no_estim.meta.Adj <- d$se.summary
LL.estim.Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_adj_mean.zinb <- data.frame(x=c("Adj mean ZINB MA: R vs SC"),y=round(c(estim.Adj.meta.4),2),
ylo=round(c(LL.estim.Adj.meta.4),2),yhi=round(c(UL.estim.Adj.meta.4),2),se=se.no_estim.meta.Adj)
logs <- c(log(ET_adj.zero.zinb),log(IQ_adj.zero.zinb)); selogs <-
c(se.Adj.ET.zinb.zero,se.Adj.IQ.zinb.zero); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.Adj.meta.4 <- round(exp(d$summary),2); se.estim.meta.Adj <- d$se.summary
LL.estim.Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_adj_zero.zinb <- data.frame(x=c("Adj zero ZINB MA: R vs
SC"),y=round(c(estim.Adj.meta.4),2),ylo=round(c(LL.estim.Adj.meta.4),2),
=round(c(UL.estim.Adj.meta.4),2),se=se.estim.meta.Adj)
##-- ET: Non-Adjusted ZIP (offset(FU))
S <- summary(m5.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data =
datos1,dist="negbin"))
estim <- c(); estim[1] <- S$coefficients$count[2,1]
estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos1)[1]){
  d <- datos1[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = d,dist="negbin"))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos1)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-61]-mean(estim.zero[-61]))^2))
ET_no_adj.mean.zinb <- exp(mean(estim))
ET_no_adj.mean.zinb_inf <- exp(mean(estim)-qt(0.975,dim(datos1)[1]-1)*desv)
ET_no_adj.mean.zinb_sup <- exp(mean(estim)+qt(0.975,dim(datos1)[1]-1)*desv)
se.no_Adj.ET.mean.zinb<- desv
ET_no_adj_mean.zinb <- data.frame(x=c("Non-Adj mean ZINB ET: R vs
SC"),y=round(c(ET_no_adj.mean.zinb),2),ylo=round(c(ET_no_adj.mean.zinb_inf),2),
yhi=round(c(ET_no_adj.mean.zinb_sup),2),se=se.no_Adj.ET.mean.zinb)
ET_no_adj.zero.zinb <- exp(mean(estim.zero))
ET_no_adj.zero.zinb_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos1)[1]-1)*desv.zero)
ET_no_adj.zero.zinb_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos1)[1]-1)*desv.zero)
se.no_Adj.ET.zinb.zero<- desv.zero
ET_no_adj_zero.zinb <- data.frame(x=c("Non-Adj zero ZINB ET: R vs
SC"),y=round(c(ET_no_adj.zero.zinb),2),ylo=round(c(ET_no_adj.zero.zinb_inf),2),
yhi=round(c(ET_no_adj.zero.zinb_sup),2),se=se.no_Adj.ET.zinb.zero)
##-- IQ: Non-Adjusted ZIP (offset(FU))
S <- summary(m5.1 <- zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data =
datos2,dist="negbin"))
estim <- c(); estim[1] <- S$coefficients$count[2,1]

```

```

estim.zero <- c(); estim.zero[1] <- S$coefficients$zero[2,1]
desv.zero <- S$coefficients$zero[2,2]
for (i in 1:dim(datos2)[1]){
  d <- datos2[-i,]
  S <- summary(zeroinfl(NC ~ as.factor(Group)+offset(Follow.up), data = d,dist="negbin"))
  estim[i] <- S$coefficients$count[2,1]
  estim.zero[i] <- S$coefficients$zero[2,1]
}
n <- dim(datos2)[1]
desv <- sqrt(((n-1)/n)*sum((estim-mean(estim))^2))
desv.zero <- sqrt(((n-1)/n)*sum((estim.zero[-61]-mean(estim.zero[-61]))^2))
IQ_no_adj.mean.zinb <- exp(mean(estim))
IQ_no_adj.mean.zinb_inf <- exp(mean(estim)-qt(0.975,dim(datos2)[1]-1)*desv)
IQ_no_adj.mean.zinb_sup <- exp(mean(estim)+qt(0.975,dim(datos2)[1]-1)*desv)
se.no_Adj.IQ.mean.zinb<- desv
IQ_no_adj.mean.zinb <- data.frame(x=c("Non-Adj mean ZINB IQ: R vs
SC"),y=round(c(IQ_no_adj.mean.zinb),2),ylo=round(c(IQ_no_adj.mean.zinb_inf),2),
yhi=round(c(IQ_no_adj.mean.zinb_sup),2),se=se.no_Adj.IQ.mean.zinb)
IQ_no_adj.zero.zinb <- exp(mean(estim.zero))
IQ_no_adj.zero.zinb_inf <- exp(mean(estim.zero)-qt(0.975,dim(datos2)[1]-1)*desv.zero)
IQ_no_adj.zero.zinb_sup <- exp(mean(estim.zero)+qt(0.975,dim(datos2)[1]-1)*desv.zero)
se.no_Adj.IQ.zinb.zero<- desv.zero
IQ_no_adj.zero.zinb <- data.frame(x=c("Non-Adj zero ZINB IQ: R vs
SC"),y=round(c(IQ_no_adj.zero.zinb),2),ylo=round(c(IQ_no_adj.zero.zinb_inf),2),
yhi=round(c(IQ_no_adj.zero.zinb_sup),2),se=se.no_Adj.IQ.zinb.zero)
#- MA
logs <- c(log(ET_no_adj.mean.zinb),log(IQ_no_adj.mean.zinb)); selogs <-
c(se.no_Adj.ET.mean.zinb,se.no_Adj.IQ.mean.zinb); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.no_Adj.meta.4 <- round(exp(d$summary),2); se.no_estim.meta.no_Adj <- d$se.summary
LL.estim.no_Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no_Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_no_adj.mean.zinb <- data.frame(x=c("Non-Adj mean ZINB MA: R vs
SC"),y=round(c(estim.no_Adj.meta.4),2),
ylo=round(c(LL.estim.no_Adj.meta.4),2),yhi=round(c(UL.estim.no_Adj.meta.4),2),se=se.no_estim.meta
.no_Adj)
logs <- c(log(ET_no_adj.zero.zinb),log(IQ_no_adj.zero.zinb)); selogs <-
c(se.no_Adj.ET.zinb.zero,se.no_Adj.IQ.zinb.zero); names <- c("ET study", "IQ study")
b <- data.frame(names,logs,selogs)
d <- meta.summaries(b$logs, b$selogs, names=b$names,method="fix", logscale=TRUE)
estim.no_Adj.meta.4 <- round(exp(d$summary),2); se.estim.meta.no_Adj <- d$se.summary
LL.estim.no_Adj.meta.4 <- round(exp(d$summary-1.96*d$se.summary),2)
UL.estim.no_Adj.meta.4 <- round(exp(d$summary+1.96*d$se.summary),2)
MA_no_adj.zero.zinb <- data.frame(x=c("Non-Adj zero ZINB MA: R vs
SC"),y=round(c(estim.no_Adj.meta.4),2),ylo=round(c(LL.estim.no_Adj.meta.4),2),
yhi=round(c(UL.estim.no_Adj.meta.4),2),se=se.estim.meta.no_Adj)

### CITATION-PER-YEAR MEAN RATIO FOREST PLOT FOR DATA COLLECTED IN 2018

# d1: NA - NB
# d2: A - NB
# d3: NA - ZIP
# d4: A - ZIP
# d5: NA - ZINB
# d6: A - ZINB

ET <- rbind(ET_no_adj_nb[2:4],
            ET_adj_nb[2:4],
            ET_no_adj_mean.zip[2:4],
            ET_adj_mean.zip[2:4],
            ET_no_adj_mean.zinb[2:4],
            ET_adj_mean.zinb[2:4])
s.ET <- rbind(ET_no_adj_nb[5],
              ET_adj_nb[5],
              ET_no_adj_mean.zip[5],
              ET_adj_mean.zip[5],
              ET_no_adj_mean.zinb[5],
              ET_adj_mean.zinb[5])

IQ <- rbind(IQ_no_adj_nb[2:4],
            IQ_adj_nb[2:4],
            IQ_no_adj_mean.zip[2:4],
            IQ_adj_mean.zip[2:4],
            IQ_no_adj_mean.zinb[2:4],
            IQ_adj_mean.zinb[2:4])
s.IQ <- rbind(IQ_no_adj_nb[5],
              IQ_adj_nb[5],
              IQ_no_adj_mean.zip[5],
              IQ_adj_mean.zip[5],
              IQ_no_adj_mean.zinb[5],
              IQ_adj_mean.zinb[5])

```

```

      IQ_adj_nb[5],
      IQ_no_adj_mean.zip[5],
      IQ_adj_mean.zip[5],
      IQ_no_adj_mean.zinb[5],
      IQ_adj_mean.zinb[5])

MA <- rbind(MA_no_adj_nb[2:4],
           MA_adj_nb[2:4],
           MA_no_adj_mean.zip[2:4],
           MA_adj_mean.zip[2:4],
           MA_no_adj_mean.zinb[2:4],
           MA_adj_mean.zinb[2:4])
s.MA <- cbind(MA_no_adj_nb[5],
             MA_adj_nb[5],
             MA_no_adj_mean.zip[5],
             MA_adj_mean.zip[5],
             MA_no_adj_mean.zinb[5],
             MA_adj_mean.zinb[5])

poin <- c(0,15)
lin <- c(1,2,2)
wid <- c(1,1,2)
pos <- c(0.5,1.25,2.25,3,4,4.75,0,6.25,7,8,8.75,9.75,10.5,0,13.5,14.25,15.25,16,17,17.75)
pos<-pos[length(pos):1]

library(extrafont)
win.graph(15,12)
par(oma=c(0,5,3,7)); par(family = "Times New Roman")
plot(NA, xlim = c(0.33,3), ylim=c(0,18),xaxt="n",yaxt="n",
     ylab="",xlab="",axes=F,log="x")
abline(v=1,lty=2)
axis(1,c(1/3.5,1/3,0.5,1,2,3,3.5),labels=F)
mtext(c("1/3","1/2","1","2","3"),las=1,side=1,at=c(1/3,0.5,1,2,3),line=0.5,cex=0.8)

par(xpd=NA)
legend(0.15,24,horiz=T,bty="n",c("Negative Binomial",
                                "Zero-Inflated Poisson",
                                "Zero-Inflated Negative Binomial"),
      lty=c(lin,0,0),lwd=c(wid,0,0),pch=c(NA,NA,NA),poin)
legend(0.5,22.5,horiz=T,bty="n",c("Non-adjusted",
                                "Adjusted"),lty=c(0,0),lwd=c(0,0),pch=c(poin))

par(xpd=FALSE)

# Joint estimation (Meta-analysis)
segments(MA$ylo[1],pos[1],MA$yhi[1],pos[1],lwd=wid[1],lty=lin[1])
points(MA$y[1],pos[1],pch=poin[1],lwd=wid[1],cex=(0.5/s.MA[,1]))
segments(MA$ylo[2],pos[2],MA$yhi[2],pos[2],lwd=wid[1],lty=lin[1])
points(MA$y[2],pos[2],pch=poin[2],cex=0.5/s.MA[,2])
segments(MA$ylo[3],pos[3],MA$yhi[3],pos[3],lwd=wid[2],lty=lin[2])
points(MA$y[3],pos[3],pch=poin[1],cex=0.5/s.MA[,3])
segments(MA$ylo[4],pos[4],MA$yhi[4],pos[4],lwd=wid[2],lty=lin[2])
points(MA$y[4],pos[4],pch=poin[2],cex=0.5/s.MA[,4])
segments(MA$ylo[5],pos[5],MA$yhi[5],pos[5],lwd=wid[3],lty=lin[3])
points(MA$y[5],pos[5],pch=poin[1],lwd=wid[2],cex=0.5/s.MA[,5])
segments(MA$ylo[6],pos[6],MA$yhi[6],pos[6],lwd=wid[3],lty=lin[3])
points(MA$y[6],pos[6],pch=poin[2],cex=0.5/s.MA[,6])
# ET
segments(ET$ylo[1],pos[8],ET$yhi[1],pos[8],lwd=wid[1],lty=lin[1])
points(ET$y[1],pos[8],pch=poin[1],lwd=wid[1],cex=0.5/s.ET[,1])
segments(ET$ylo[2],pos[9],ET$yhi[2],pos[9],lwd=wid[1],lty=lin[1])
points(ET$y[2],pos[9],pch=poin[2],cex=0.5/s.ET[,2])
segments(ET$ylo[3],pos[10],ET$yhi[3],pos[10],lwd=wid[2],lty=lin[2])
points(ET$y[3],pos[10],pch=poin[1],cex=0.5/s.ET[,3])
segments(ET$ylo[4],pos[11],ET$yhi[4],pos[11],lwd=wid[2],lty=lin[2])
points(ET$y[4],pos[11],pch=poin[2],cex=0.5/s.ET[,4])
segments(ET$ylo[5],pos[12],ET$yhi[5],pos[12],lwd=wid[3],lty=lin[3])
points(ET$y[5],pos[12],pch=poin[1],lwd=wid[2],cex=0.5/s.ET[,5])
segments(ET$ylo[6],pos[13],ET$yhi[6],pos[13],lwd=wid[3],lty=lin[3])
points(ET$y[6],pos[13],pch=poin[2],cex=0.5/s.ET[,6])
# IQ
segments(IQ$ylo[1],pos[15],IQ$yhi[1],pos[15],lwd=wid[1],lty=lin[1])
points(IQ$y[1],pos[15],pch=poin[1],lwd=wid[1],cex=0.5/s.IQ[,1])
segments(IQ$ylo[2],pos[16],IQ$yhi[2],pos[16],lwd=wid[1],lty=lin[1])
points(IQ$y[2],pos[16],pch=poin[2],cex=0.5/s.IQ[,2])
segments(IQ$ylo[3],pos[17],IQ$yhi[3],pos[17],lwd=wid[2],lty=lin[2])
points(IQ$y[3],pos[17],pch=poin[1],cex=0.5/s.IQ[,3])
segments(IQ$ylo[4],pos[18],IQ$yhi[4],pos[18],lwd=wid[2],lty=lin[2])

```

```

points(IQ$y[4],pos[18],pch=point[2],cex=0.5/s.IQ[4,])
segments(IQ$ylo[5],pos[19],IQ$yhi[5],pos[19],lwd=wid[3],lty=lin[3])
points(IQ$y[5],pos[19],pch=point[1],lwd=wid[2],cex=0.5/s.IQ[5,])
segments(IQ$ylo[6],pos[20],IQ$yhi[6],pos[20],lwd=wid[3],lty=lin[3])
points(IQ$y[6],pos[20],pch=point[2],cex=0.5/s.IQ[6,])
mtext("Mean Ratio (95% CI)",las=2,side=2,at=19.5,line=-32.5,font=2)
mtext("Joint MA",las=2,side=2,at=pos[3],line=2.7,font=2,cex=1.2)
mtext("estimated effect",las=2,side=2,at=pos[4],line=0,font=2,cex=1.2)
mtext("ET study",las=2,side=2,at=pos[10],line=3,font=2)
mtext("(n=41 vs 51)",las=2,side=2,at=pos[11],line=2)
mtext("IQ study",las=2,side=2,at=pos[17],line=3,font=2)
mtext("(n=27 vs 24)",las=2,side=2,at=pos[18],line=2)
mtext(paste(MA$y,"(",MA$ylo,",",MA$yhi,")",sep=""),side=4,las=2,at=pos[1:6],line=1.5,cex=1)
mtext(paste(ET$y,"(",ET$ylo,",",ET$yhi,")",sep=""),side=4,las=2,at=pos[8:13],line=1.5,cex=0.9)
mtext(paste(IQ$y,"(",IQ$ylo,",",IQ$yhi,")",sep=""),side=4,las=2,at=pos[15:20],line=1.5,cex=0.9)
mtext("Higher in intervention",las=2,side=2,at=-2.3,line=-27,cex=0.9)
mtext("Higher in control",las=2,side=2,at=-2.3,line=-4,cex=0.9)

### OR FOREST PLOT FOR DATA COLLECTED IN 2018

# d3: NA - ZIP
# d4: A - ZIP
# d5: NA - ZINB
# d6: A - ZINB

ET <- rbind(ET_no_adj_zero.zip[2:4],
            ET_adj_zero.zip[2:4],
            ET_no_adj_zero.zinb[2:4],
            ET_adj_zero.zinb[2:4])
s.ET <- rbind(ET_no_adj_zero.zip[5],
              ET_adj_zero.zip[5],
              ET_no_adj_zero.zinb[5],
              ET_adj_zero.zinb[5])

IQ <- rbind(IQ_no_adj_zero.zip[2:4],
            IQ_adj_zero.zip[2:4],
            IQ_no_adj_zero.zinb[2:4],
            IQ_adj_zero.zinb[2:4])
s.IQ <- rbind(IQ_no_adj_zero.zip[5],
              IQ_adj_zero.zip[5],
              IQ_no_adj_zero.zinb[5],
              IQ_adj_zero.zinb[5])

MA <- rbind(MA_no_adj_zero.zip[2:4],
            MA_adj_zero.zip[2:4],
            MA_no_adj_zero.zinb[2:4],
            MA_adj_zero.zinb[2:4])
s.MA <- rbind(MA_no_adj_zero.zip[5],
              MA_adj_zero.zip[5],
              MA_no_adj_zero.zinb[5],
              MA_adj_zero.zinb[5])

poin <- c(0,15)
lin <- c(1,1,1)
wid <- c(1,1,1)
pos <- c(0.5,1.25,2.25,3,5,5.75,6.75,7.5,11.5,12.25,13.25,14)
pos<-pos[length(pos):1]

win.graph(18,14)
par(oma=c(0,4,0,9)); par(family="Times New Roman")
plot(NA, xlim = c(1/100000,100000), ylim=c(0,15),xaxt="n",yaxt="n",
     ylab="",xlab="",axes=F,log="x")
abline(v=1,lty=2)
axis(1,c(1/1000000,1/100000,1/10000,1/1000,1/100,1/10,0,10,100,1000,10000,100000,1000000),labels=
F)
mtext(c("1/10^5","1/10^4","1/10^3","1/10^2","1/10","0","10","10^2","10^3","10^4","10^5"),las=1,si
de=1,at=c(1/100000,1/10000,1/1000,1/100,1/10,0,10,100,1000,10000,100000),line=0.5,cex=0.8)
par(xpd=NA)
legend(0.000004,18.5,horiz=T,bty="n",c("Zero-Inflated Poisson",
"Zero-Inflated Negative
Binomial"),lty=1,col=c("black","darkgrey"))
legend(0.005,17.5,horiz=T,bty="n",c("Non-adjusted","Adjusted"),lty=c(0,0),lwd=c(0,0),pch=c(poin))
par(xpd=FALSE)

# Pooled estimation (Meta-analysis)
arrows(1,pos[1],270000,pos[1],lwd=wid[2],lty=lin[2],length=0.12)
arrows(1,pos[1],1/270000,pos[1],lwd=wid[2],lty=lin[2],length=0.12)

```

```

points(MA$y[1],pos[1],pch=poin[1],cex=10/s.MA[1,])
arrows(1,pos[2],270000,pos[2],lwd=wid[2],lty=lin[2],length=0.12)
arrows(1,pos[2],1/270000,pos[2],lwd=wid[2],lty=lin[2],length=0.12)
points(MA$y[2],pos[2],pch=poin[2],cex=10/s.MA[2,])
arrows(MA$ylo[3],pos[3],270000,pos[3],lwd=wid[3],lty=lin[3],length=0.12,col="darkgrey")
points(MA$y[3],pos[3],pch=poin[1],lwd=wid[3],cex=10/s.MA[3,],col="darkgrey")
arrows(1,pos[4],270000,pos[4],lwd=wid[3],lty=lin[3],length=0.12,col="darkgrey")
arrows(1,pos[4],1/270000,pos[4],lwd=wid[3],lty=lin[3],length=0.12,col="darkgrey")
points(MA$y[4],pos[4],pch=poin[2],cex=10/s.MA[4,],col="darkgrey")
# ET
arrows(1/270000,pos[5],270000,pos[5],lwd=wid[1],lty=lin[1],length=0.12)
arrows(270000,pos[5],1/270000,pos[5],lwd=wid[1],lty=lin[1],length=0.12)
points(ET$y[1],pos[5],pch=poin[1],lwd=wid[1],cex=10/s.ET[1,])
arrows(1,pos[6],270000,pos[6],lwd=wid[1],lty=lin[1],length=0.12)
arrows(1,pos[6],1/270000,pos[6],lwd=wid[1],lty=lin[1],length=0.12)
points(ET$y[2],pos[6],pch=poin[2],cex=10/s.ET[2,])
arrows(1,pos[7],270000,pos[7],lwd=wid[2],lty=lin[2],length=0.12,col="darkgrey")
arrows(1,pos[7],1/270000,pos[7],lwd=wid[2],lty=lin[2],length=0.12,col="darkgrey")
points(ET$y[3],pos[7],pch=poin[1],cex=10/s.ET[3,],col="darkgrey")
arrows(1/270000,pos[8],270000,pos[8],lwd=wid[1],lty=lin[1],length=0.12,col="darkgrey")
arrows(270000,pos[8],1/270000,pos[8],lwd=wid[1],lty=lin[1],length=0.12,col="darkgrey")
points(ET$y[4],pos[8],pch=poin[2],cex=10/s.ET[4,],col="darkgrey")
# IQ
arrows(1/270000,pos[9],270000,pos[9],lwd=wid[1],lty=lin[1],length=0.12)
arrows(270000,pos[9],1/270000,pos[9],lwd=wid[1],lty=lin[1],length=0.12)
points(IQ$y[1],pos[9],pch=poin[1],lwd=wid[1],cex=10/s.IQ[1,])
arrows(1/270000,pos[10],270000,pos[10],lwd=wid[1],lty=lin[1],length=0.12)
arrows(270000,pos[10],1/270000,pos[10],lwd=wid[1],lty=lin[1],length=0.12)
points(IQ$y[2],pos[10],pch=poin[2],cex=10/s.IQ[2,])
arrows(1/270000,pos[11],270000,pos[11],lwd=wid[1],lty=lin[1],length=0.12,col="darkgrey")
arrows(270000,pos[11],1/270000,pos[11],lwd=wid[1],lty=lin[1],length=0.12,col="darkgrey")
points(IQ$y[3],pos[11],pch=poin[1],cex=10/s.IQ[3,],col="darkgrey")
arrows(1/270000,pos[12],270000,pos[12],lwd=wid[2],lty=lin[2],length=0.12,col="darkgrey")
arrows(270000,pos[12],1/270000,pos[12],lwd=wid[2],lty=lin[2],length=0.12,col="darkgrey")
points(IQ$y[4],pos[12],pch=poin[2],cex=10/s.IQ[4,],col="darkgrey")
mtext("OR (95% CI)",las=2,side=2,at=15.5,line=-31.5,font=2)
mtext("Join MA",las=2,side=2,at=pos[2],line=4,font=2)
mtext("estimated effect",las=2,side=2,at=pos[2]-0.5,line=1.5,font=2)
mtext("ET study",las=2,side=2,at=pos[6],line=4,font=2,cex=0.9)
mtext("(n=41 vs 51)",las=2,side=2,at=pos[6]-0.5,line=3,cex=0.9)
mtext("IQ study",las=2,side=2,at=pos[10],line=4,font=2,cex=0.9)
mtext("(n=27 vs 24)",las=2,side=2,at=pos[10]-0.5,line=3,cex=0.9)
mtext(paste(MA$y,"(",MA$ylo,"",
",c("6.17*10^10","1.33*10^12","7.17*10^9","4.16*10^10"),"),sep=""),side=4,las=2,at=pos[1:4],line
=1.5,cex=1)
mtext(paste(ET$y,"(",ET$ylo,"",
",c("4.73*10^14","8.78*10^16","6.37*10^12","2.17*10^12"),"),sep=""),side=4,las=2,at=pos[5:8],lin
e=1.5,cex=0.9)
mtext(paste(IQ$y,"(",IQ$ylo,"",
",c("3.44*10^15","3.35*10^15","1.23*10^12","2.43*10^15"),"),sep=""),side=4,las=2,at=pos[9:12],li
ne=1.5,cex=0.9)
mtext("Higher in intervention",las=2,side=2,at=-2.1,line=-26,cex=0.9)
mtext("Higher in control",las=2,side=2,at=-2.3,line=-4,cex=0.9)

```

```
#####
##-- GRÀFIC CDFs
#####
rm(list=ls())
library(xmeta)
library(pscl)

##-- Udated data in 31 December 2018
datos <- read.csv('data_2018.csv',header=TRUE,sep=";")
datos$Follow.up <- as.numeric((as.Date("31/12/2018", "%d/%m/%Y") -
as.Date(datos$Publication.date, "%d/%m/%Y"))/365)
datos1 <- datos[datos$Study=="ET",]

library(pscl)
library(boot)
library(fitdistrplus) # fits distributions using maximum likelihood
library(gamlss) # defines pdf, cdf of ZIP

##### ET STUDY
round(prop.table(table(datos1$NC))*100,2)
ks1 <- ecdf(datos1$NC)
## Poisson
fit.pois <- fitdist(datos1$NC, 'pois')
round(prop.table(table(rpois(fit.pois$n, fit.pois$estimate)))*100,2)
ks2 <- ecdf(rpois(fit.pois$n, fit.pois$estimate))
## Negative Binomial
fit.nb <- fitdist(datos1$NC, 'nbinom')
round(prop.table(table(rnegbin(fit.nb$n, fit.nb$estimate["mu"], fit.nb$estimate["size"])))*100,2)
## ZIP
S <- summary(m1 <- zeroinfl(NC ~ 1, data = datos1))
mu_C = exp(S$coefficients$count[1]); sigma_C = S$coefficients$count[2]
zero_C = inv.logit(S$coefficients$zero[1])
set.seed(1234)
fit_zip = fitdist(datos1$NC, 'ZIP', start = list(mu = mu_C, sigma = zero_C)
, lower=c(-Inf, 0.000001), upper=c(Inf, 1))
round(prop.table(table(rZIP(fit_zip$n, fit_zip$estimate["mu"], fit_zip$estimate["sigma"])))*100,2)
## ZINB
S <- summary(zeroinfl(NC ~ 1, data = datos1, dist="negbin", link="logit"))
mu_C = exp(S$coefficients$count[1]); zero_C = inv.logit(S$coefficients$zero[1]); sigma_C =
S$theta
set.seed(4321)
fit.zinb <- rZINBI(length(datos1$NC), mu=mu_C, sigma=sigma_C, nu=zero_C)
h<-round(prop.table(table(fit.zinb))*100,2)
h<-hist(fit.zinb)

## Plot CDFs for ET study
library(extrafont)
windows()
par(family="Times New Roman")
cdfcomp(list(fit.nb, fit_zip, fit.pois), lwd=2, addlegend = F, xlab="Number of
citations", fitlty=c(1,1,1))
lines(ecdf(fit.zinb), verticals=T, lwd=2, col="orange", lty=1, pch=NA)
lines(ecdf(fit_zip), verticals=T, lwd=2, col="blue", lty=1, pch=NA)
legend("bottomright", c("Sample", "Poisson", "Zero-Inflated Poisson", "Negative Binomial", "Zero-
Inflated Negative Binomial", "")
, col=c("black", "blue", "green", "red", "orange", "white"), lty=1, lwd=2, bty="n")

windowsFonts("Times New Roman" = windowsFont("Times New Roman"))
win.graph(9,9)
par(family="Times New Roman", mfrow=c(2,1))
denscomp(list(fit_zip, fit.pois), lwd=2, ylim=c(0,0.2), xlim=c(0,15), addlegend = F, xlab="Number of
citations", fitlty=1, datacol = c("black", "green", "blue"))
legend("topright", c("Sample", "Poisson", "Zero-Inflated Poisson")
, col=c("black", "green", "red"), lty=1, lwd=2, bty="n")
denscomp(list(fit.nb), lwd=2, ylim=c(0,0.2), xlim=c(0,15), addlegend = F, xlab="Number of
citations", fitlty=1, main="")
d <- density(h, lty=1, col="orange", lwd=2)
t<- tapply(d$y, round(d$x), mean) [5:21]
segments(0.15:15.15, 0, 0.15:15.15, t, lty=1, col="green", lwd=2)
legend("topright", c("Sample", "Negative Binomial", "Zero-Inflated Negative Binomial")
, col=c("black", "green", "red"), lty=1, lwd=2, bty="n")
```



```
#####
##
## CODE TO COMPUTE K-S STATISTIC ESTIMATION WITH 95% CI
##
##-- Description:
## - random generation of samples with Pois, Q-P, NB, ZIP, ZINB
## - ks.test() funtion of R to estimate K-S statistic
## - 10000 simulations to obtain a 95% CI
##
#####

rm(list=ls())
library(rmeta)
library(pscl)

##-- Udated data in 31 December 2018
datos <- read.csv('data_2018.csv',header=TRUE,sep=";")
datos$Follow.up <- as.numeric((as.Date("31/12/2018", "%d/%m/%Y")-
as.Date(datos$Publication.date, "%d/%m/%Y"))/365)
datos1 <- datos[datos$Study=="ET",]
datos2 <- datos[datos$Study=="IQ" & (datos$Group=="R" | datos$Group=="SC"),]

nsim <- 10000
n <- dim(datos1)[1]
##-- Poisson
set.seed(1234)
ks.pois <- c()
for(i in 1:nsim){d <- rpois(n, lambda = mean(datos1$NC)); ks.pois[i] <-
ks.test(datos1$NC,d)$statistic}
KS.POIS <- median(ks.pois); LL.KS.POIS <- quantile(ks.pois,0.025); UL.KS.POIS <-
quantile(ks.pois,0.975)
##-- Negative-binomial
library(fitdistrplus)
set.seed(1234)
fit.nb <- fitdist(datos1$NC, 'nbinom')
ks.nb <- c()
for(i in 1:nsim){d <- rnegbin(n, mu = fit.nb$estimate[2], theta = fit.nb$estimate[1]); ks.nb[i]
<- ks.test(datos1$NC,d)$statistic}
KS.NB <- median(ks.nb); LL.KS.NB <- quantile(ks.nb,0.025); UL.KS.NB <- quantile(ks.nb,0.975)
##-- Zero-Inflated Poisson
set.seed(1234)
library(pscl); library(boot); library(gamlss)
S <- summary(m1 <- zeroinfl(NC ~ 1, data = datos1))
mu = exp(S$coefficients$count[1])
sigma = inv.logit(S$coefficients$zero[1])
ks.zip <- c()
for(i in 1:nsim){d <- rZIP(n, mu, sigma); ks.zip[i] <- ks.test(datos1$NC,d)$statistic}
KS.ZIP <- median(ks.zip); LL.KS.ZIP <- quantile(ks.zip,0.025);UL.KS.ZIP <- quantile(ks.zip,0.975)
##-- Zero-Inflated Negative Binomial
set.seed(1234)
library(pscl)
model.zinb <- zeroinfl(NC ~ 1, data = datos1,dist="negbin")
S <- summary(model.zinb); k <- S$theta ; lambda <- exp(S$coefficients$count[1]);
omega = inv.logit(S$coefficients$zero[1]); ks.zinb <- c()
for(i in 1:nsim){d <- rZINBI(n, k, lambda, omega); ks.zinb[i] <- ks.test(datos1$NC,d)$statistic}
KS.ZINB <- median(ks.zinb); LL.KS.ZINB <- quantile(ks.zinb,0.025); UL.KS.ZINB <-
quantile(ks.zinb,0.975)
##-- Taula
taula <- rbind(cbind(KS.POIS,LL.KS.POIS,UL.KS.POIS),
              cbind(KS.NB,LL.KS.NB,UL.KS.NB),
              cbind(KS.ZIP,LL.KS.ZIP,UL.KS.ZIP),
              cbind(KS.ZINB,LL.KS.ZINB,UL.KS.ZINB))
rownames(taula) <- c("Poisson","Negative Binomial", "Zero-Inflated Poisson","Zero-Infalted
Negative Binomial")
colnames(taula) <- c("K-S","95%LL","95%UL")
round(taula,2)

nsim <- 10000
n <- dim(datos2)[1]
##-- Poisson
set.seed(1234)
ks.pois <- c()
for(i in 1:nsim){d <- rpois(n, lambda = mean(datos2$NC)); ks.pois[i] <-
ks.test(datos2$NC,d)$statistic}
KS.POIS <- median(ks.pois); LL.KS.POIS <- quantile(ks.pois,0.025); UL.KS.POIS <-
quantile(ks.pois,0.975)
##-- Negative-binomial
```

```

library(fitdistrplus)
set.seed(1234)
fit.nb <- fitdist(datos2$NC, 'nbinom')
ks.nb <- c()
for(i in 1:nsim){d <- rnegbin(n, mu = fit.nb$estimate[2], theta = fit.nb$estimate[1]); ks.nb[i]
<- ks.test(datos2$NC,d)$statistic}
KS.NB <- median(ks.nb); LL.KS.NB <- quantile(ks.nb,0.025); UL.KS.NB <- quantile(ks.nb,0.975)
##-- Zero-Inflated Poisson
set.seed(1234)
library(pscl); library(boot); library(gamlss)
S <- summary(m1 <- zeroinfl(NC ~ 1, data = datos2))
mu = exp(S$coefficients$count[1])
sigma = inv.logit(S$coefficients$zero[1])
ks.zip <- c()
for(i in 1:nsim){d <- rZIP(n, mu, sigma); ks.zip[i] <- ks.test(datos2$NC,d)$statistic}
KS.ZIP <- median(ks.zip); LL.KS.ZIP <- quantile(ks.zip,0.025);UL.KS.ZIP <- quantile(ks.zip,0.975)
##-- Zero-Inflated Negative Binomial
set.seed(1234)
library(pscl)
model.zinb <- zeroinfl(NC ~ 1, data = datos2,dist="negbin")
S <- summary(model.zinb); k <- S$theta ;lambda <- exp(S$coefficients$count[1]);
omega = inv.logit(S$coefficients$zero[1]); ks.zinb <- c()
for(i in 1:nsim){d <- rZINBI(n, k,lambda, omega); ks.zinb[i] <- ks.test(datos2$NC,d)$statistic}
KS.ZINB <- median(ks.zinb); LL.KS.ZINB <- quantile(ks.zinb,0.025); UL.KS.ZINB <-
quantile(ks.zinb,0.975)
##-- Taula
taula <- rbind(cbind(KS.POIS,LL.KS.POIS,UL.KS.POIS),
              cbind(KS.NB,LL.KS.NB,UL.KS.NB),
              cbind(KS.ZIP,LL.KS.ZIP,UL.KS.ZIP),
              cbind(KS.ZINB,LL.KS.ZINB,UL.KS.ZINB))
rownames(taula) <- c("Poisson","Negative Binomial", "Zero-Inflated Poisson","Zero-Infalted
Negative Binomial")
colnames(taula) <- c("K-S","95%LL","95%UL")
round(taula,2)

```

```
#####
##
## RECURRENT EVENTS ANALYSIS
##
#####
rm(list=ls())

plt.rcParams["font.family"] = "Times New Roman"

dat <- read.csv("BD_RecurrentEvents_2020.csv",header=T,sep=";")
dat$Publication.date<-as.Date(dat$Publication.date,"%d/%m/%Y")
dat$sp1[dat$Previous.study=="ET"] <- (as.Date("2020-12-31","%Y-%m-%d") -
dat$Publication.date[dat$Previous.study=="ET"])/365
dat.ET <- dat[dat$Previous.study=="ET",]
dat.ET <- dat.ET[order(dat.ET$Article.ID,dat.ET$t.stop.2020),]
id <- dat.ET[duplicated(dat.ET$Article.ID)==F,c("Article.ID","Publication.date","p1")]
id_ordenat <- id[order(id$Publication.date),]
library(extrafont)
##-- Histograms of the number of citations, by study group and by trial
windowsFonts(A = windowsFont("Times New Roman"))
windows(12,11)
par(family = "Times New Roman")
par(oma=c(0,0,0,0))
plot(NA,xaxt="n",yaxt="n",
      ylim=c(1,92),xlim=c(0,12.5),lwd=0.5,ylab="Article",xlab="Year",main="",family="A")
axis(1,at=0:12,labels=c("2008","2009","2010","2011","2012","2013","2014","2015","2016","2017","20
18","2019","2020"),family="A")
axis(2,at=1:92,labels=id_ordenat$Article.ID,las=2,family="A",cex.axis=0.4,tick=F)
for(i in 1:dim(id_ordenat)[1]){

segments(id_ordenat$sp1[i],i,max(dat.ET$t.stop.2020[dat.ET$Article.ID==id_ordenat$Article.ID[i]])+
id_ordenat$sp1[i],i,
col=ifelse(dat.ET$Group[dat.ET$Article.ID==id_ordenat$Article.ID[i]]=="R","orange","blue"),lwd=1)
}
for(i in 1:dim(id_ordenat)[1]){
  x <- dat.ET$t.stop.2020[dat.ET$Article.ID==id_ordenat$Article.ID[i]]+id_ordenat$sp1[i]
  points(x[-length(x)],rep(i,length(x)-1),pch=15,cex=0.7)
}
for(i in 1:dim(id_ordenat)[1]){
  x <- dat.ET$t.stop.2020[dat.ET$Article.ID==id_ordenat$Article.ID[i]]+id_ordenat$sp1[i]
  points(x[length(x)],i,pch=1,cex=0.7)
}
par(xpd=NA)
legend(4,105,c("R","SC"), col=c("orange","blue"),pch=15,horiz = T,bty = "n")

## MCF AMB LA LLIBRERIA REDA DE R per BD (data 20/07/2020)
library(reda)
dat_MCF <-
data.frame(ID=dat$Article.ID,time=as.numeric(dat$t.stop.2020),Event=as.numeric(dat$Event),
           Study=dat$Previous.study,Study.type=dat$Study.type,Group=dat$Group,
           GS_B=dat$Baseline.Goodman.Scale..likert.5.,GS_F=dat$Final.Goodman.Scale..Likert.5.)
dat_MCF_ET <- dat_MCF[dat_MCF$Study=="ET",]
mod1 <- mcf(Survr(ID, time, Event)~1,data=dat_MCF_ET,variance="LawlessNadeau")
windows()
library(ggplot2)
par(family = "Times New Roman")
plot(mod1,conf.int=TRUE,mark.time=T,addOrigin=T,col="grey3",ylim=c(0,6),family="A") +
ggplot2::xlab("Years") +
ggplot2::ylab("Mean cumulative number of citations") + ggplot2::ggtitle("") + #Sample Mean
Cumulative Function for ET study
ggplot2::ylim(c(0,6)) + ggplot2::xlim(c(0,11)) + ggplot2::theme_bw() +
ggplot2::theme(text=element_text(family="Times New Roman", size=12))
# ET segons Grup
library(ggplot2)
dat_MCF_ET$Group2 <- relevel(as.factor(dat_MCF_ET$Group),ref="SC")
mod1 <- mcf(Survr(ID, time, Event)~Group2,data=dat_MCF_ET,variance="LawlessNadeau")
windows()
plot(m<-mcfDiff(mod1),col="navy",lwd=2) + ggplot2::xlab("Years") + ggplot2::theme_bw() +
ggplot2::ylab("MCF difference between SC and R groups") + ggplot2::ggtitle("Sample MCF
differences between groups for ET study") +
ggplot2::xlim(c(0,12)) + ggplot2::ylim(c(-4,4))+ geom_hline(yintercept=0)+
geom_hline(yintercept=0) + ggplot2::theme(text=element_text(family="Times New Roman", size=12))

## FRAILTY MODEL WITH frailtySurv package (semi-parametric)
rm(list=ls())
library(frailtySurv)
```

```
head(dat)
dat_ET <- dat[dat$Previous.study=="ET",]
dat_ET$Group2 <- as.factor(as.character(dat_ET$Group))
vfit <- survfit(Surv(t.start.2018, t.stop.2018+0.0001, Event) ~ Group2, data=dat_ET,
id=Article.ID)
windows()
plot(vfit, cumhaz=FALSE, col=1:2, conf.times=c(1, 3, 5, 7, 9),
      xlab="Years since publication", ylab="Cumulative hazard")
legend("topright", col=1:2, levels(dat_ET$Group2), lty=1)
```