Living Meta-Review Generation for Social Scientists: An Interface and A Case Study on Human Cooperation

Inès Blin[1,2,*], Ilaria Tiddi[1], Giuliana Spadaro[1] and Annette ten Teije[1]

[1]Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2]Sony Computer Science Laboratories-Paris, Paris, France

Abstract

In the social sciences, researchers formulate a hypothesis, conduct background research and run a meta-analysis to test that hypothesis synthesizing a large body of evidence. Results of the meta-analysis are published in a *meta-review*. Yet this process is manual, demanding significant effort and expert knowledge, and results in static PDF formats. Furthermore, social scientists struggle in keeping the reviews up to date due to the increasing number of empirical studies. A *living, automated meta-review generation system* would be the solution to ease this process. We present an interface to select starting hypotheses and to create living meta-reviews in HTML format automatically, with a case study on human cooperation. This will enable to assist social scientists in their daily tasks, and we believe that the benefits of this work will be broadly relevant across the social sciences.

Keywords

living meta-review, automatic meta-review generation, interface, demonstration paper

1. Introduction

Let us imagine social scientists investigating the impact of gender on cooperative behaviors. This example follows one from the literature [1] and will be our running example. Social scientists must search for and aggregate findings from studies measuring cooperative behavior in relation to gender. These data exist but are often dispersed across individual studies or domain-specific reviews in non-machine readable formats, making their access complicated. Social scientists typically conduct a meta-analysis [2], the gold standard for synthesizing research from studies. A meta-analysis systematically reviews literature using statistical techniques to combine and compare results from related studies [3]. Traditional steps of a meta-analysis include: hypothesis formulation, literature search, study selection, dependent variables choice, meta-analysis model selection, result analysis, and interpretation [1, 4, 5]. However, this process is manual, demanding significant effort and expert knowledge.

Recent work [6, 7, 8] has contributed to a knowledge graph (KG) paradigm based on an interlinked and formal description of research publications [9]. This would alleviate the background research and foster FAIR principles [10] such as transparency and reproducibility. CS-KG [6] gathers entities and claims from the computer science domain, ORKG [7] curates semantic scholarly knowledge from research papers, and the COoperation DAtabank (CoDa) [8] describes findings on human cooperation in social science. Building KGs from published articles can take time, but automation is accelerating this process. Integrating annotation into the publication process would enable faster data access.

Such datasets, and the research platforms built upon them [11], can be queried to retrieve studies that tested relationships between variables, perform on-demand meta-analyses, estimate publication bias and statistical power analyses to inform future studies with unprecedented ease. Yet, results of these meta-analyses manually gathered in published meta-review are mostly in textual, static PDF

EKAW 2024: EKAW 2024 Workshops, Tutorials, Posters and Demos, 24th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), November 26-28, 2024, Amsterdam, The Netherlands.

*Corresponding author.
i.blin@vu.nl (I. Blin); i.tiddi@vu.nl (I. Tiddi); g.spadaro@vu.nl (G. Spadaro); annette.ten.teije@vu.nl (A. t. Teije)
0000-0003-0956-9466 (I. Blin); 0000-0001-7116-9338 (I. Tiddi); 0000-0002-5706-5898 (G. Spadaro); 0000-0002-9771-8822 (A. t. Teije)

formats.[1] In the R Shiny[2] interface[3] [11] built on top of CoDa, one can run many meta-analyses, but cannot retrieve summary meta-reviews. Data providers thus struggle to keep meta-reviews up to date, which in turn hinders the scientific understanding of rapidly evolving fields. For instance, between 2018 and 2022, approximately 11,000 articles were published on cooperation in economic games [8].

A *living, automated meta-review generation* system summarizing meta-analyses would be the solution to accommodate this challenge. By enabling researchers to interact with the data, such a tool would enhance the utility of meta-analytic knowledge. The meta-analyses would be easily findable in an open, standard format, fostering reproducibility and credibility. Additionally, as the living meta-analyses are dynamically built on structured datasets and automatically updated with data changes, they would provide the most current information to guide future research.

We present an interface to create *living meta-reviews* automatically built from data queried on demand. The meta-review in HTML format becomes dynamic with the regeneration of the HTML each time a hypothesis is modified. We present a use case in human cooperation. This will aid social scientists in summarizing empirical evidence, reducing human efforts at all levels of the data management process like publication or querying. We believe the benefits of this work will extend beyond our case study and will be broadly relevant across the social sciences. To the best of our knowledge, we are the first ones to present such a novel interface. We make our code to build the interface openly available.[4]

2. Interface

Overview. Our interface allows users to select a hypothesis and to generate a meta-review in HTML format automatically. Studies comparing two groups are retrieved from the *hypothesis* and refined from *inclusion criteria*. The *analytic strategy* defines the statistical model to run the meta-analysis, and *control variables* are displayed in the meta-review. Lastly, *custom content* can be added before *generating and displaying the meta-review*. We propose a step-by-step overview of the interface's functionalities, and we make a demonstration video of the interface publicly available.[5]

1. Select a hypothesis. This is the first step where the user can choose a hypothesis based on the comparison of two different groups. This enables fetching studies for the meta-analysis.

We define terms used by social scientists [12]. An effect size (*es*) [13] measures the relationship strength between two variables in a study. A dependent variable (*dv*) is the outcome that is measured in a study to assess the effect of changes in the independent variable(s). Our case study investigates human cooperation, using CoDa as the KG backend resource. It contains around 3,000 studies from the social and behavioural sciences, annotated with cooperation-related features. Generic independent variables (*givs*) are broad factors categorized to observe their effect on a *dv*. Specific independent variables (*siv*) are well-defined variables within a study, categorized to determine their effect on a *dv*. Specific independent variable values (*sivvs*) are the value assigned to the *siv*. *es*, *dv*, *giv*, *siv* and *sivv* are assumed from CoDa. The left part of Figure 1 provides an example of the structure of CoDa: the paper `id:ENG00006` includes the study `id:ENG00006_1` that reports the effect `id:ENG00006_1.1.1.2.d` comparing the treatments `id:00006_1.1.2` and `id:ENG00006_1.1.1`, hereafter denoted as T1 and T2. T1 and T2 have different characteristics for the `cp:gender` property, `female` and `male`.

We consulted with two experts to design meaningful hypotheses. While we lack formal evaluation, the experts agreed that comparing only identical or related independent variables is meaningful. Furthermore, user studies we conducted with 5 domain experts on the same templated hypotheses did not raise any objection. The main technical constraint we impose is thus that the *sivvs* to be compared be instances of the same *siv*, which is in the ontology. The right part of Figure 1 shows the

[1] Static examples of such meta-reviews can be found here: https://cooperationdatabank.org/living-reviews/
[2] https://shiny.posit.co
[3] https://app.cooperationdatabank.org
[4] https://github.com/InesBlin/coda_meta_review/tree/main
[5] https://bit.ly/ekaw-pd-2024-meta-review-generation

hypothesis we chose as a running example for this paper. CoDa contains three *dvs*: `cooperation`, its subcategory `contributions`, and `withdrawals`, with 9,456, 9,948, and 1,774 reported effects, respectively. Since `withdrawals` had fewer studies, we grouped `contributions` with `cooperation`, considering `cooperation` as the sole *dv*. The user can choose *giv*, *siv* and *sivv* in this step, and *es* in Step 4.

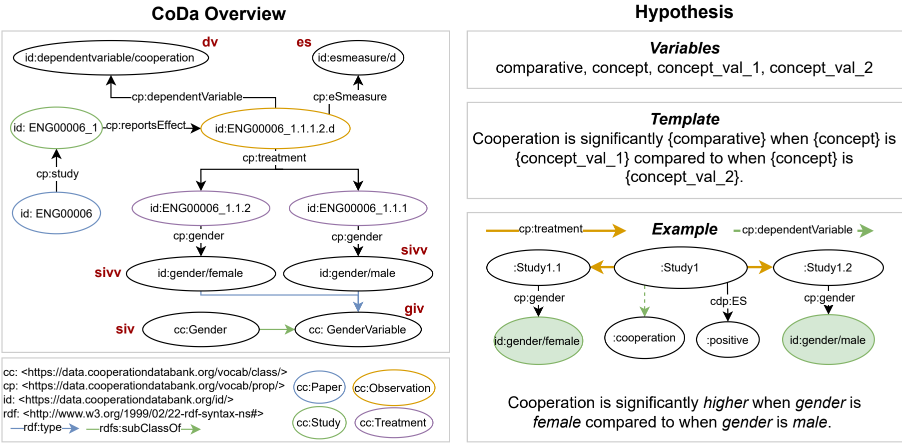

Figure 1: Visual overview of the CoDa dataset (left) and of one example hypothesis to be selected (right).

The options dynamically update based on the current selection, ensuring they remain in sync with CoDa. Figure 2 shows the interface with our running example, from which the hypothesis is derived.

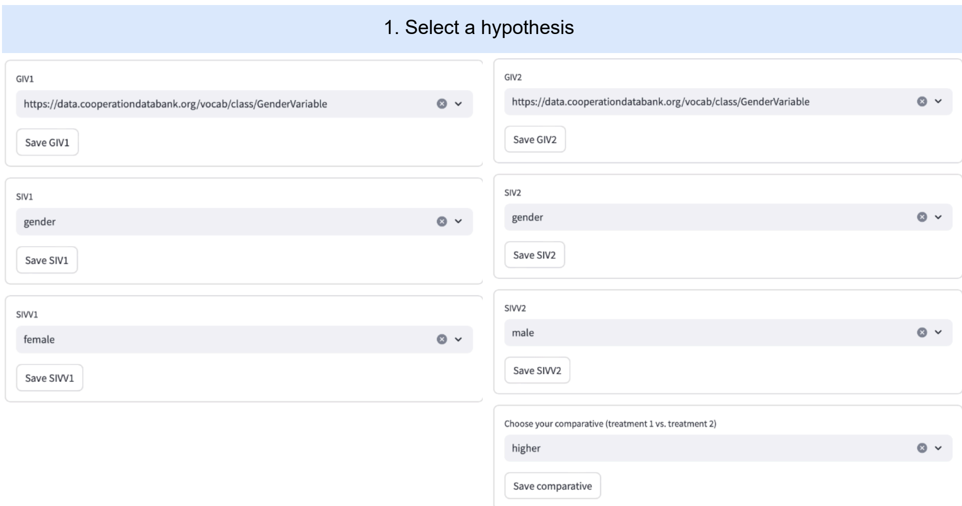

Figure 2: Overview of the hypothesis selection step in the interface. The user first selects the first and second group's characteristics, and then the cooperation comparative ("higher" or "lower") between these two groups.

2. Inclusion criteria. This step refines study filters for the meta-analysis. There are four types of criteria: (1) metadata, (2) quantitative (3) sample based on participants characteristics, and (4) study. The left part of Figure 3 shows an example, where the user only keeps studies from published articles.

3. Control variables. This step enables to choose additional variables to be described and displayed in the meta-review, such as `maleProportion` and `discussion` on the right part of Figure 3 in the interface.

4. Analytic strategy. This step concerns the statistical model to run on the selected studies to test the hypothesis for the meta-analysis, like a simple restricted maximum likelihood (REML) model [14] with Cohen's standardized mean difference [15] as effect size on the left part of Figure 4 in the interface.

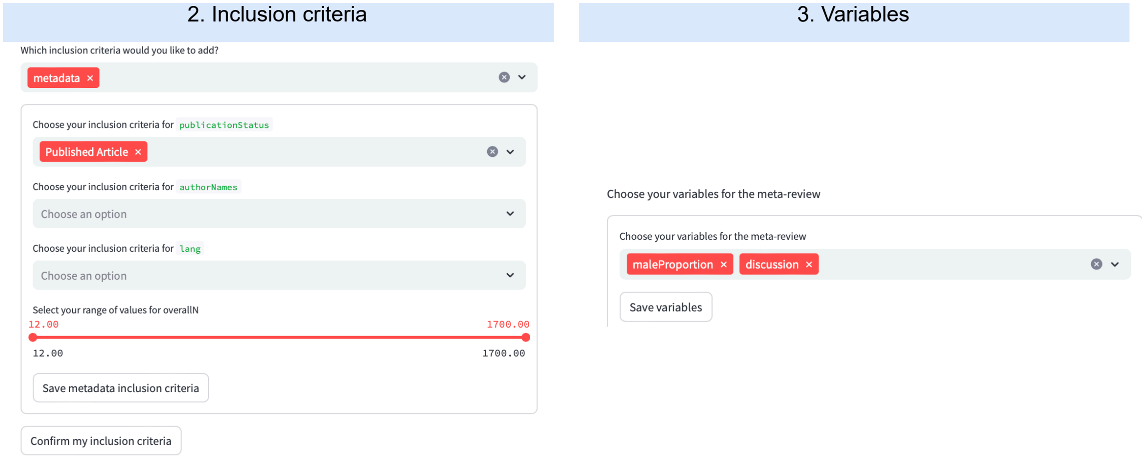

Figure 3: Overview of the inclusion criteria (left) and variable selection (right) steps in the interface.

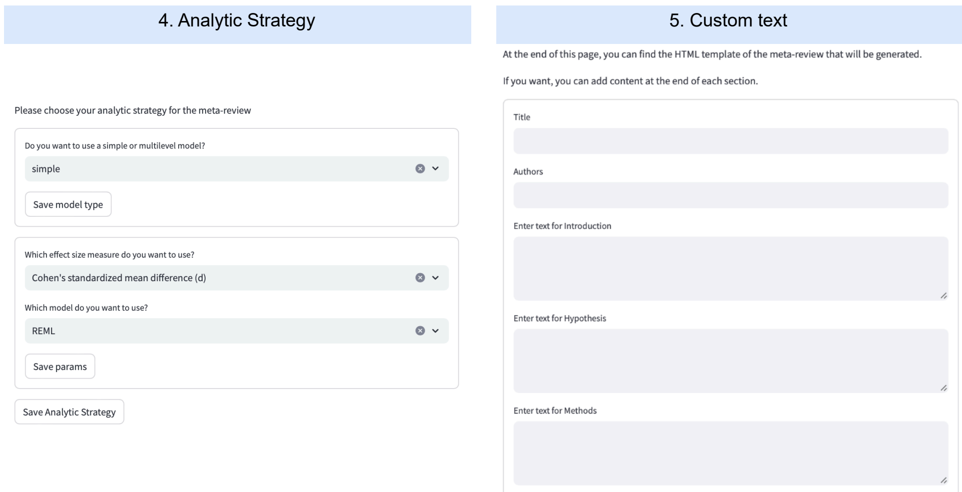

Figure 4: Overview of the analytic strategy (left) and custom text (right) steps in the interface.

5. Custom text. This step is the possibility for experts to provide more descriptive information in the form of custom text, as shown on the right part of Figure 4, and to be added to the meta-review.

6. Generate and display the meta-review. The meta-review is based on an HTML template that was carefully designed with domain experts, and is generated based on previous criteria. One example of output can be seen in Figure 5.

Technical details. Regarding **data storage and management**, CoDa is hosted by Triply[6] in the form of 7 Named KGs. To make the interface more efficient, we used SPARQL queries, available in the code, to create *.csv* backend files. For the **frontend/user interface**, we used Streamlit,[7] and plotly[8] for most visualisations. The interface is in the app folder of the repository. The core functions that are called in the interface are in the `src` folder. For the **backend**, we used Python,[9] and called the R package `metafor`[10] for the meta-analyses. The interface is not yet deployed live.

[6]https://odissei.triply.cc/coda/databank
[7]https://streamlit.io
[8]https://plotly.com
[9]https://www.python.org
[10]https://cran.r-project.org/web/packages/metafor/index.html

Custom Title

John Doe et al.

1. Introduction

In this meta-analysis, we test whether women cooperate more than men.

2. Hypothesis

The objective of this meta-review is to explore a starting hypothesis on differences in cooperation, and investigate whether the CoDa databank supports this hypothesis or not.

Hypothesis explored: Cooperation is significantly higher when gender is female compared to when gender is male.

Treatment 1: GIV = https://data.cooperationdatabank.org/vocab/class/GenderVariable, SIV = gender, SIVV = female

Treatment 2: GIV = https://data.cooperationdatabank.org/vocab/class/GenderVariable, SIV = gender, SIVV = male

3. Methods

3.1. Inclusion Criteria

Studies included in the analyses were retrieved through the Cooperation Databank (CoDa) [1]. Details of the systematic search that led to the annotation of the studies are reported in full in Spadaro et al. [1].

For studies to be included in the meta-analysis, the following criteria had to be fulfiled:

- metadata: {'publicationStatus': ['Published Article'], 'overallN': [12.0, 1700.0]}

These data and criteria resulted in the inclusion of 180 effect sizes extracted from 120 studies conducted across 27 industrialized societies.

Study Provenance | Study Year | Study Data

3.2. Coding of effect sizes

To capture standardized mean difference, we used Cohen's d.

3.3. Coding of variables

Show 10 entries Search:

node	label	description
discussion	Discussion	Whether communication was allowed between participants in the game.
maleProportion	Proportion of males	Proportion of male participants in the sample of a study (after exclusion of participants, when that information was reported).

Showing 1 to 2 of 2 entries Previous 1 Next

3.4. Analytic strategy

All analyses were conducted using the metafor package (Viechtbauer, 2010) in R (RCoreTeam, 2019).

To test the hypothesis that "Cooperation is significantly higher when gender is female compared to when gender is male.", we first estimated the overall effect size using a mixed effects model with REML estimation, along with both the 95% confidence interval (CI) and the 90% prediction interval.

We then considered the variation in the effect size distribution by using several indicators of heterogeneity of variance (T, T^2, and I^2).

4. Results

Run a Meta-analysis

Raw R Output | Structured R Output

The meta-analysis is based on $k = 180$ effects. Findings revealed a small overall meta-analytic effect of $d = -0.02$ (95% CI [-0.071, 0.028], 90% PI [-0.071, 0.028]).

This result indicates that " there is no significant difference in cooperation when comparing studies where gender is female and studies where gender is male."

There is heterogeneity in the effect size distribution ($T = 0.238$, $T^2 = 0.057$) and a substantial amount of this variation can be explained by between-study differences ($I^2 = 60.689$).

Moreover, the overall effect size distribution contained more variation than would be expected by chance, $Q(179) = 441.68$, $p = 0.0$.

References

[1] Spadaro, Giuliana and Tiddi, Ilaria and Columbus, Simon and Jin, Shuxian and Ten Teije, Annette and CoDa Team and Balliet, Daniel (2022). The Cooperation Databank: Machine-Readable Science Accelerates Research Synthesis.

Figure 5: Overview of the HTML meta-review in the interface. The selected hypothesis was: "Cooperation is significantly *higher* when *gender* is *female* compared to when *gender* is *male*".

3. Conclusion and Future Work

We present an interface to create living meta-reviews based on hypotheses comparing two groups with different characteristics. This will aid social scientists in the process of summarizing empirical evidence on their research questions, reducing the human efforts at all levels of the data management process, such as publication or querying. Ultimately, the goal is to index, aggregate and centralize meta-reviews to avoid duplication. Quality will be ensured through peer review or expert validation. The interface is now limited to one type of hypothesis only, and there is currently no user evaluation. We plan to integrate more complex hypotheses and to extend meta-reviews templates to improve the tool for social scientists. We also plan to deploy the interface online, and to conduct user studies to better assess the usefulness of the tool. We are also working on automatically suggesting hypotheses with various AI methods, as well as user studies to evaluate the hypotheses.

Acknowledgments

This work was partly funded by the European MUHAI project, grant no. 951846, and by the XS NWO Project, grant no. 406.XS.04.118. We thank our reviewers for constructive comments.

References

[1] G. Spadaro, S. Jin, D. Balliet, Gender differences in cooperation across 20 societies: a meta-analysis, Philosophical Transactions of the Royal Society B: Biological Sciences 378 (2022) 20210438. URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2021.0438. doi:10.1098/rstb.2021.0438, publisher: Royal Society.

[2] A. F. H. Ortiz, E. C. Camacho, J. C. Rojas, T. C. Camacho, S. Z. Guevara, N. T. R. Cuenca, A. V. Perdomo, V. D. C. Herazo, R. G. Malo, A Practical Guide to Perform a Systematic Literature Review and Meta-analysis, Principles and Practice of Clinical Research 7 (2021) 47–57. URL: https://www.journal.ppcr.org/index.php/ppcrjournal/article/view/175. doi:10.21801/ppcrj.2021.74.6, number: 4.

[3] G. V. Glass, Primary, secondary, and meta-analysis of research, Educational researcher 5 (1976) 3–8.

[4] M. Yuan, G. Spadaro, S. Jin, J. Wu, Y. Kou, P. A. M. Van Lange, D. Balliet, Did cooperation among strangers decline in the United States? A cross-temporal meta-analysis of social dilemmas (1956–2017), Psychological Bulletin 148 (2022) 129–157. doi:10.1037/bul0000363, place: US Publisher: American Psychological Association.

[5] L. Bucher, T. Burgard, U. S. Tran, G. M. Prinz, M. Bosnjak, M. Voracek, Keeping Meta-Analyses Alive and Well: A Tutorial on Implementing and Using Community-Augmented Meta-Analyses in PsychOpen CAMA, Advances in Methods and Practices in Psychological Science 6 (2023) 25152459231197611. URL: https://doi.org/10.1177/25152459231197611. doi:10.1177/25152459231197611, publisher: SAGE Publications Inc.

[6] D. Dessí, F. Osborne, D. Reforgiato Recupero, D. Buscaldi, E. Motta, CS-KG: A Large-Scale Knowledge Graph of Research Entities and Claims in Computer Science, in: U. Sattler, A. Hogan, M. Keet, V. Presutti, J. P. A. Almeida, H. Takeda, P. Monnin, G. Pirrò, C. d'Amato (Eds.), The Semantic Web – ISWC 2022, Springer International Publishing, Cham, 2022, pp. 678–696. doi:10.1007/978-3-031-19433-7_39.

[7] M. Stocker, A. Oelen, M. Y. Jaradeh, M. Haris, O. A. Oghli, G. Heidari, H. Hussein, A.-L. Lorenz, S. Kabenamualu, K. E. Farfar, M. Prinz, O. Karras, J. D'Souza, L. Vogt, S. Auer, FAIR scientific information with the Open Research Knowledge Graph, FAIR Connect 1 (2023) 19–21. URL: https://content.iospress.com/articles/fair-connect/fc221513. doi:10.3233/FC-221513, publisher: IOS Press.

[8] G. Spadaro, I. Tiddi, S. Columbus, S. Jin, A. Ten Teije, C. Team, D. Balliet, The cooperation databank: machine-readable science accelerates research synthesis, Perspectives on Psychological Science 17 (2022) 1472–1489.

[9] A. A. Salatino, Y. Bu, Y. Ding, A. Horvát, Y. Huang, M. Liu, P. Manghi, A. Mannocci, F. Osborne, D. Romero, D. Sacharidis, M. Teplitskiy, T. Vergoulis, F. Xia, Y. Zhai, 3rd International Workshop on Scientific Knowledge Representation, Discovery, and Assessment (Sci-K 2023), in: Companion Proceedings of the ACM Web Conference 2023, WWW '23 Companion, Association for Computing Machinery, New York, NY, USA, 2023, p. 783. URL: https://doi.org/10.1145/3543873.3589740. doi:10.1145/3543873.3589740.

[10] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. URL: https://www.nature.com/articles/sdata201618. doi:10.1038/sdata.2016.18, publisher: Nature Publishing Group.

[11] I. Tiddi, D. Balliet, A. ten Teije, Fostering scientific meta-analyses with knowledge graphs: a

case-study, in: The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings 17, Springer, 2020, pp. 287–303.
[12] A. Bryman, Social research methods, Oxford university press, 2016.
[13] K. Kelley, K. J. Preacher, On effect size., Psychological methods 17 (2012) 137.
[14] Y. Dodge, The Oxford dictionary of statistical terms, Oxford University Press, USA, 2003.
[15] J. Cohen, Statistical power analysis for the behavioral sciences, routledge, 2013.