Community-Augmented Meta-Analyses (CAMAs) in Psychology

Potentials and Current Systems

Tanja Burgard , Michael Bošnjak , and Robert Studtrucker

Leibniz Institute for Psychology (ZPID), Trier, Germany

Abstract: The limits of static snapshot meta-analyses and the relevance of reproducibility and data accessibility for cumulative meta-analytic research are outlined. A publication format to meet these requirements is presented: Community-augmented meta-analyses (CAMA). We give an overview of existing systems implementing this approach and compare these in terms of scope, technical implementation, data collection and augmentation, data curation, tools available for analysis, and methodological flexibility.

Keywords: replicability, meta-analysis, cumulative research, open data, repository

Typically, meta-analyses are published exclusively as static snapshots, depicting the evidence in a specific area up to a certain point in time. Moreover, in psychology, published meta-analyses rarely meet common reporting standards, such as the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which was conceptualized more than a decade ago, or the more recently suggested MARS (Meta-analysis reporting standards) (Lakens et al., 2017). This practice leads to serious limitations with regard to the reusability of meta-analytic data and the currency of evidence.

The first problem often encountered by researchers is the lack of information to replicate the results of a meta-analysis. As a response to this problem, Lakens et al. (2016) argue for open meta-analytic data to make meta-analyses dynamic and reproducible. This is important for several reasons. First, having open access meta-analytic data would enable researchers the possibility of examining the sensitivity of the results to subjective decisions that were made in the original process of synthesizing the data, such as the underlying inclusion criteria, statistical models, or use of moderators. Second, an open access meta-analyses register would enable the application of new statistical procedures to existing data and allow testing the effects that these have on the meta-analytic results. Third, open access to existing meta-analyses provides other researchers

with special research questions the opportunity to use subsets of the preexisting meta-analytic data (Bergmann et al., 2018).

The second problem of static snapshot meta-analyses is the fact that they are only valid for a specific cut-off date (Créquit et al., 2016). Without additional electronic material, a meta-analysis represents the cumulative evidence on a research question up to a certain point in time and may quickly become outdated as soon as new findings from primary studies are published or new methodological or statistical procedures are developed (Shojania et al., 2007). If the data are no longer accessible, the time-consuming process of conducting a meta-analysis must start from the beginning.

To facilitate and simplify cumulative research and to strengthen the evidence, for example, if practical challenges call for clear recommendations and decisions, we need to think about how to effectively publish our meta-analyses to make knowledge production more efficient. The key challenges for the publication of meta-analyses, therefore, are to make the preexisting research reproducible and to allow the updating of meta-analyses by reusing the information that has been collected up to the point of the most recent meta-analysis. In the following, these challenges will be discussed and the requirements for a publication format that enables reproducible and dynamic meta-analyses will be derived.

Challenges and Requirements for Meta-Analyses

Reproducibility and Replicability

Scientific findings can be validated on different levels (Stanley et al., 2018). Reproducibility means to produce exactly the same results with the same data and analyses. To validate findings at this level, accessibility of data and analysis code are sufficient. On the next level, we aim for replicability. When the same results and conclusions are obtained as those found in the original study using a new random sample and following the reported procedures, we can report a successful replication of results. Moreover, we distinguish between direct and conceptual replications (Zwaan et al., 2018). For direct replication, all critical facets of the study design in the original study have to be captured. A conceptual replication allows some differences in the study procedures. If findings are replicated independently of unmeasured factors in the original study, such as, for instance, sample characteristics or the country where the study took place, a finding can be considered as generalizable.

Replicability is already a problem at the level of individual studies. Reasons for failed replications can be found in every phase of the research cycle. If the findings are already known, researchers may propose hypotheses after the fact (a phenomenon known as HARKing – hypothesizing after the results are known; Kerr, 1998), leading to significant results deriving from their dataset. Another questionable research practice used to obtain significant results from the data is *p*-hacking (Simonsohn et al., 2014). Of the 64 studies they investigated, Banks et al. (2016) identified evidence of questionable research practices in 91%.

Apart from the questionable validity, individual studies often suffer from low statistical power (Fraley & Vazire, 2014), so that it is unlikely to find effects – especially small ones – even if they actually exist. Poor study quality and relevant differences in study design can also lead to different study results. Finally, published studies may not be representative of all studies, as significant results are published more often (Ioannidis, 2008). By chance, studies will provide meaningful results from time to time, even if the effects do not really exist. However, these results are often not replicable.

In meta-analyses, we expect a higher validity of the results due to the stronger evidence base and the heterogeneity in study designs and samples captured with a meta-analysis (Borenstein et al., 2009, p. 9). However, due to the frequency of questionable research practices in individual studies, the risk of bias of these should be accounted for in the first place (Hohn et al., 2019). Above this, when conducting a meta-analysis, there are a number

of subjective decisions and sources of error threatening its validity (Cooper, 2017, p. 318).

Table 1 provides an overview of potential errors in each step of a meta-analysis. There might be several plausible alternatives for some decisions such as, for example, model specifications or treatment of missing data. However, the decisions must be justified and clearly reported to allow replication of the meta-analysis and also to investigate the impact of these decisions on the results by means of sensitivity analyses.

For meta-analyses, the four principles (open access, open methodology, open data, and source) of the open science movement advocated by Kraker et al. (2011) may be applied to overcome the challenge of making meta-analyses reproducible. Based on these principles, we derive the following requirements:

1. The transparent documentation of all steps and decisions along the meta-analytic process as presented in Table 1 enables the assessment of possible biases.
2. Common standards for interoperable and usable open data and scripts allow the verification of the results of a review. Subjective decisions may be modified, and new procedures may be applied with minimal effort to check the robustness of the results.

Updating and Cumulative Evidence

Static snapshot meta-analyses may quickly become outdated if they lack reusability to be expandable. Regular updates of meta-analyses are necessary. For example, Cochrane reviews should be updated every 2 years (Shojania et al., 2007) and Campbell reviews within 5 years (Lakens et al., 2016). Créquit et al. (2016) examined the proportion of available evidence on lung cancer not covered by systematic reviews between 2009 and 2015 with the finding that, in all cases, at least 40% of treatments were missing.

For systematic reviews, an update is defined as a new edition of a published review. It can include new data, new methods, or new analyses. An update is recommended if the topic is still relevant and new methods or new studies have emerged that could potentially change the findings of the original review (Garner et al., 2016).

Shojania et al. (2007) define signals of relevant evidence changes to warrant the update of reviews. These signals are changes in statistical significance, a relevant relative change in effect magnitude, new information on the clinical relevance of a review, or the emergence of new approaches not considered previously. For 100 reviews, the time between the publication and the occurrence of a signal for updating is measured and the median survival time of a meta-analysis in their analysis is 5.5 years. Within 2 years,

Table 1. Selected potential sources of error in the conduction of a meta-analytic process

Step of the meta-analysis according to Cooper (2017)	Sources of error	Threats to validity
Formulating the problem	Poorly defined constructs and relationships (e.g., Baer et al., 2019)	Questionable construct validity of measures
Searching the literature	Studies found in the literature search may not correctly represent the relevant population of studies (e.g., Pietschnig et al., 2010; Kepes & McDaniel, 2015)	Publication bias
Gathering information	Incorrect information retrieval resulting in misrepresentations in the data (e.g., London, 2016)	Unreliability in coding (inter- and intrapersonal), biased effect size sampling (favoring one direction of findings)
Quality appraisal	Evaluation approach not exhaustive concerning study design characteristics, nonexplicit weighting scheme (e.g., Hohn et al., 2019)	Biased exclusion of studies, biased weighting of studies
Synthesis methods and analysis	Nonweighting of effect sizes, unjustified model specifications, for example, in case of heterogeneous or dependent effect sizes (e.g., Voracek et al., 2019)	Biased estimates and inferences
Interpretation of results	Inaccurate treatment of missing data, no consideration of confounded moderator effects and heterogeneity in samples and study setting (e.g., Tran et al., 2014)	Biased estimates, overgeneralization of findings to contexts not represented in the meta-analysis

almost one-fourth of the reviews were already outdated (Shojania et al., 2007). As the number of publications is continuously growing (Bastian et al., 2010), we can expect the survival times of reviews to become even shorter.

Meta-analyses can also reveal research gaps by providing an overview of potential moderators or moderator combinations not yet sufficiently studied. In the case of Zhu et al. (2014), previous meta-analyses on the effect of thiazolidinedione treatment on the risk of fractures had mainly focused on postmenopausal women. New evidence provided the opportunity to study gender as a potential moderator of the effect, and it turned out that increased risk of fractures was only detected for women.

The ongoing accumulation of evidence informs researchers about the latest findings in a specific research area, for example, when the results are robust enough to no longer justify further research investment, at least without taking into account existing results and perhaps specific research gaps. A systematic review of cumulative meta-analyses (Clarke et al., 2014) reports many illustrating examples, speaking for the high relevance of cumulative research to enable more informed decisions and at the same time a more efficient distribution of research funds and efforts.

As requirements to overcome the challenge of updating meta-analyses, we can thus derive:

1. There is a need for infrastructures that are able to monitor the currentness and validity of meta-analytic evidence and to provide and apply decision rules for the necessity of updates.
2. Open access to data and metadata provides pre-existing research a usable and sustainable future. Extracted

metadata and coding can be used to update a meta-analysis or even to conduct another meta-analysis on a similar subject with an overlap in the relevant literature.

3. Accumulating science and keeping evidence updated is a cooperative task and the participation in this task has to be supported and incentivized, for example, as proof of achievement instead of, or in addition to, the classical single publication.

Community-Augmented Meta-Analysis (CAMA) as a Publication Format

Openly available and regularly updated meta-analyses support the efficiency of science. Researchers can get a quick overview of a research field, can use the latest evidence for power analyses and study planning, and may make use of curated information and data to identify research gaps, as understudied moderator variables. As a solution for comprehensive, dynamic, and up-to-date evidence synthesis, Créquit et al. (2016) call for living systematic reviews, that is high-quality online summaries, that are continuously updated. Similarly, Haddaway (2018) proposes open synthesis.

Actually, a concept for a publication format for meta-analyses that meets these requirements already exists. There are slightly different forms that have been suggested for this meta-analytical concept including living (Elliott et al.,

Figure 1. The basic structure of a CAMA

2017), dynamic (Bergmann et al., 2018), or cloud-based meta-analysis (Bosco et al., 2015). Braver et al. (2014) describe an approach called continuously cumulating meta-analysis (CCMA) to incorporate and evaluate new replication attempts to existing meta-analyses. In our conception, we use the term community-augmented meta-analysis, CAMA for short (Tsuji, Bergmann, & Cristia, 2014). A CAMA is a combination of an open repository for meta-analytic data and an interface offering meta-analytic analysis tools.

The core of a CAMA, as shown in Figure 1, is the data repository, where meta-analytic data contributions from researchers in specific research areas are stored. It serves as a dynamic resource and can be used and augmented by the research community to keep the state of research updated and accumulate knowledge continuously. Tools to replicate and modify analyses with these data are accessible via an open web-based platform, usually encompassing a graphical user interface (GUI). For example, examining moderator effects beyond the analyses presented in the original meta-analysis may be conducted. The available evidence from the meta-analyses archived in a CAMA can also be used to improve study planning. Estimates of the expected size of an effect can serve as input for power analyses. The examination of possible relevant moderators can help to identify research gaps and guide the design of new studies (Tsuji et al., 2014).

Overview of Existing Systems Implementing CAMA in Psychology

There are already several systems and initiatives in psychology aiming at developing an infrastructure for the continuous curation and updating of meta-analytic evidence

and, thereby, fulfilling the call to make meta-analyses reproducible and dynamic. In the following, five of these systems are reviewed and compared. These systems have been identified by conference meetings (metaBUS and MetaLab were presented at the Research Synthesis Conferences 2018 and 2019), and by successive searches for similar systems. However, the selection is not exhaustive. There are other CAMA systems outside psychology and the life sciences (e.g., MitiGate: <https://mitigate.ibers.aber.ac.uk/>), and systems aiming for open meta-analyses, but providing less information and guidance for users, thereby rendering them less adequate for comparison purposes (e.g., openMetaAnalysis: <https://openmetaanalysis.ocpu.io/home/www/>).

A project located in the domain of management and applied psychology is metaBUS. It is based on a hierarchical taxonomy of the field and provides a database consisting of correlations between clearly defined concepts within this taxonomy (Bosco et al., 2020). MetaBUS is a cloud-based platform and search engine providing access to more than 1.1 million curated findings from over 14,000 articles published in applied psychology journals since 1980 (<https://www.metaBUS.org>). It relies on the RStudio Shiny architecture for the GUI (Bosco et al., 2015) and the R statistics package metafor (Viechtbauer, 2010) for the meta-analytic calculations and visualizations.

To collaboratively collect and curate meta-analyses in the fields of early language acquisition and cognitive development, MetaLab offers a shiny webapp to reproduce meta-analyses and visualizations conducted with the statistical software R (Tsuji et al., 2017). Unlike the approach of metaBUS to retrieve single correlations, the data in MetaLab is organized in single meta-analyses, each focusing on the experimental evidence of one specific phenomenon (Bergmann et al., 2018). These meta-analyses are modified and improved collaboratively over time. At the moment (March 2020), MetaLab consists of 22 meta-analyses with information from 477 papers reporting on about 1,804 effect sizes (<http://metallab.stanford.edu/>).

Primarily located in the fields of cognitive and social psychology, the crowdsourced platform Curate Science (<https://curatescience.org>) allows the permanent curation of findings by the psychological research community. The design of the platform is guided by a unified curation framework enabling a systematic evaluation of empirical research along four dimensions: the transparency of methods and data, the reproducibility of the results by repeating the same procedures on the original data, the robustness of the results to different analytic decisions, and the replicability of effects in new samples under similar conditions (LeBel et al., 2018).

In the field of life sciences and health, Cochrane is piloting a project called Living Systematic Reviews (LSRs,

Table 2. Characteristics of selected CAMA systems in psychology and the life sciences

Domain	MetaBUS	MetaLab	Curate Science	Cochrane LSR	PsychOpen CAMA
Thematic scope	Management and applied psychology	Language acquisition and cognitive development	Psychology in general	Life sciences and health	Psychology in general
Data structure and effect sizes	Hierarchical taxonomy of concepts, over 1.1 million correlations	Single meta-analyses with over 1,800 effect sizes	Curated findings allowing evaluation (1,127 re-plications of 168 effects)	Published reports only	Single meta-analyses with effect sizes of any type
Analysis engine	metafor	metafor	metafor	None	metafor
Data collection and curation	Semi-automatic data collection and curation	Template sheets for standardization, responsible curators	Crowdsourcing and peer review to curate findings in evidence collections	Searches, screening, and updating by review author on a regular basis	Template sheets, synergy effects with related products, user accounts
User interface	R shiny webapp	R shiny webapp	Django/React (early beta, only open to small group of researchers)	None	PHP web application
Possible specifications	Multilevel, filter options, trim-and-fill parameters	Multilevel, moderator, and filter options	Filter replications by study characteristics	None	Multilevel, filter, moderator, effect size
Functionalities and tools	Flexible querying, funnel, violin plot, model output	Funnel, forest, violin, power plot, model output	Search and evaluate findings, forest plot	Follow updates on study website	Funnel, forest, power plot, model output, EGM
Export functionalities	Modification and download of query results	Dataset download as csv or excel file	None	None	Link to PsychNotebook for further analyses

Synnot et al, 2017), suggesting continuous updating for reviews with a high priority for health decision making, low certainty in the existing evidence, or a high likelihood of emerging evidence affecting the conclusions. An LSR is a review that is continually updated and incorporates new evidence immediately (Elliott et al., 2017). Cochrane LSRs and corresponding updates are published in the Cochrane database of systematic reviews (<https://www.cochranelibrary.com>).

At the Leibniz Institute for Psychology (ZPID), PsychOpen CAMA is currently under development with a first version becoming available in 2021. This service aims to serve the psychological research community as a whole by covering different psychological domains and meta-analyses on diverse effect sizes and study types. The approach for data storage and curation is similar to the one of MetaLab. Single meta-analyses can be published via the platform to become accessible to and expandable by the community. Instead of using an R shiny architecture for the GUI, PsychOpen CAMA relies on a PHP web application with an OpenCPU server for the R calculations. This improves the scalability of the web application according to the number of users, which is of special relevance for a service provided by a research infrastructure institute covering a broad scope of potential research domains, possibly reaching more users than rather narrowly specified applications targeted to a small research community.

Comparison of Data Collection, Augmentation, and Curation Approaches

The systems differ in terms of how data are collected and stored, augmented, and curated. As the data repository is the basis of a CAMA system, we will compare the systems previously introduced in terms of data administration. Table 2 sums up the central results of the comparison between the systems in terms of both data administration and, as discussed in the next section, data analyses.

The effect size of interest in metaBUS is correlations. On average, an empirical article contains 75 zero-order correlations, many of which would be overlooked during a traditional literature search. These correlations are collected by a semi-automated matrix extraction protocol. Trained coders supervise this process and additionally classify each variable according to the hierarchical taxonomy of variables and constructs in applied psychology. For each variable, further attributes, such as its reliability and response rate, are coded. The metaBUS database is constantly growing, but it relies exclusively on recruited, trained, and paid coders, as crowdsourcing efforts have not paid off yet due to the difficulty to motivate and train potential collaborators (Bosco et al., 2020).

As mentioned above, MetaLab is organized in single meta-analyses. The founders of the project have defined a general structure of potentially relevant meta-analyses.

Thus, the core parts of each meta-analysis are standardized. Templates and tutorials to explain how data have to be extracted and coded using this standardized structure are provided to guide external contributors when updating or adding meta-analyses to MetaLab (Tsuji et al., 2017). To guarantee the quality of data added to a meta-analysis, there is a responsible curator for each dataset. The standardized data in MetaLab allow the computation of common effect size measures as odds ratios or standardized mean differences. Meta-analyses in MetaLab are organized following a multilevel approach. Data usually originate from experimental studies sometimes reporting multiple effect sizes in one study and perhaps various studies in one paper. As effect sizes within a study and studies within a paper are usually more similar than effect sizes between studies or papers, the shared variance has to be considered to provide unbiased estimations (Bergmann et al., 2018).

The approach of Curate Science mainly relies on crowdsourcing. It provides a decentralized platform for the research community to curate and evaluates each other's findings. To facilitate this, Curate Science offers various features. A labeling system allows researchers to indicate compliance with reporting standards for their studies to curate transparency. The curation of reproducibility and robustness is supported by uploading corresponding reanalyses. Finally, replicability is curated by allowing the addition of replications to preexisting collections of published effects and by enabling researchers to create new evidence collections. To ensure the quality of this crowdsourced data collection, new replications added to evidence collections are reviewed by other users or editors (LeBel et al., 2018).

Research syntheses published as Cochrane reviews can be suggested for continuous updating due to special relevance. In this case, they follow clearly defined update scenarios. Searches and screening for LSRs are conducted on a regular basis (e.g., monthly). If no new data is found, only the search date is reported. If new evidence is found, the decision must be made about whether it should be integrated immediately or at a later date. In the case of immediate updating, data is extracted, analyses are rerun, and the review is republished (Elliott et al., 2017). Because this task is time-intensive for the individuals responsible for the LSR (typically the authors), there are aspirations to crowdsource and automatize microtasks for LSRs in the future. Searches may be continuously monitored by LSR specific filters at bibliographic databases, registries, and repositories. Thus, notifications may automatically be pushed in case of new potentially relevant studies. Their eligibility is assessed either by machine learning classifiers alone or complemented by crowdsourced efforts. Automation technologies for data extraction, synthesis, and reporting are still rudimentary (Thomas et al., 2017), and curation

systems enabling the research community to maintain up-to-date evidence might be a better solution so far.

The meta-analytic data for PsychOpen CAMA is stored in PsychArchives, ZPID's archive for digital research objects in psychology. To update meta-analyses or to add completely new meta-analyses to PsychOpen CAMA, ZPID will ideally rely on synergy effects with its own related services and products. Research data from primary studies in PsychArchives can be used to update corresponding meta-analyses in CAMA. Alternatively, the results of studies or even complete meta-analyses preregistered at ZPID, as well as data collected in PsychLab will be used to extend the database for PsychOpen CAMA. As MetaLab, the template for data extraction assumes a multilevel structure and aims at standardizing data from different meta-analyses. In the future, user accounts should also serve data augmentation by giving users the possibility to edit data, for example, by adding new moderators or new studies. The suggestions made by users within their own accounts, however, have to be peer-reviewed before meta-analyses are updated accordingly.

Comparison of Available Analysis Tools

At the side of the user, the presented CAMA systems differ largely in the meta-analytic functionalities and the flexibility of the tools offered via the GUI. As the GUI is crucial for the accessibility of the meta-analyses to the interested users without expertise in meta-analysis, we will focus on the provided tools of the CAMA systems in the following.

The core functionality of metaBUS is the flexible querying via exact letter strings or taxonomic classifiers. There are two report modes. For the targeted search, two search terms are specified. Moreover, dependence in effect sizes may be considered, parameters for the trim-and-fill analysis can be specified, as well as the ranges of sample size, publication year, and the correlations (Bosco et al., 2015). An instant meta-analysis over all relevant bivariate relations and the corresponding metadata are returned. Users may refine their query for example via filtering by reliability or by checking the exact operationalizations of the concepts and if necessary, exclude individual entries. The newly developed exploratory search only requires one taxonomic node and instantly reports all meta-analyses with all other taxonomic nodes via an interactive plot (Bosco et al., 2020).

MetaLab offers meta-analytic modeling options, as the use of multilevel grouping, empirical Bayes estimations, and the use of selected moderator variables. Basic visualization tools, such as violin, forest, and funnel plots are available. Furthermore, prospective power analyses informed by the meta-analytic effect size of a given meta-analysis may

be conducted to improve study planning. A simulation tool allows observing potential outcomes depending on key parameters of studies. Next to these basic tools available through a point-and-click interface, advanced users may also download the complete meta-analytic datasets and conduct their own analyses (Tsuji et al., 2017).

Curate Science essentially enables users to search for studies and evaluate findings based on characteristics related to transparency, reproducibility, robustness, and replicability. It provides an overview of the evidence on published and perhaps controversial effects. It also allows the meta-analysis of replications selected on the basis of study characteristics such as methodological similarities or preregistration status. Forest plots and meta-analytic estimations are then reported for the effect of interest (LeBel et al., 2018).

In their report on pilot LSR, Millard et al. (2018) sum up the processes and publication outputs from eight LSRs maintained during the pilot period. Depending on the amount of evidence published during this period, searches were conducted in time frames ranging from daily to once every three months. Updates were communicated on a regular basis to the readers via the study websites. For all but one study, new evidence was found during the pilot period. Only one LSR was completely republished. An interactive GUI, such as those used by metaBUS and MetaLab, is not yet available for Cochrane LSRs.

PsychOpen CAMA provides a user interface with basic meta-analysis tools, such as forest plots, funnel plots, and meta-analytic estimation. For these analyses, different effect sizes are available, dependency in effect sizes can be considered using a multilevel approach, and potentially relevant moderator variables can be chosen to be included in the model. Tools designed to inform about study planning decisions, such as evidence gap maps and power analyses, will also be included. Moreover, CAMA will be linked to PsychNotebook, a cloud-based electronic lab notebook for statistical analyses. Advanced users interested in applications that go beyond those directly available in CAMA may use the meta-analytic datasets within PsychNotebook.

To conclude, metaBUS, MetaLab, and PsychOpen CAMA address the open science principles to a great extent by providing data download and open analyses. The risk of bias of meta-analyses is also minimized by giving users the opportunity to filter included study results, including unpublished studies, add unconsidered moderator variables, and modify model specifications. Curate Science and LSR have no data export functionalities. Thus, relevant dimensions of the risk of bias, such as unjustified model specifications and unconsidered moderator effects, remain an issue, as the opportunities for open data and open analysis are not given.

Future Challenges for CAMAs

With a growing number of publications, efficient accumulation and synthesis of knowledge become the key to making scientific results usable and valid thus enabling more informed decisions. The survival time of the synthesized evidence in static meta-analyses, in many cases, is short. To keep this information up-to-date, the publication of meta-analyses in a format allowing the reuse of data already collected and an easy avenue to verify, update and modify meta-analyses is beneficial for the research community and the public.

A solution to enable dynamic and reusable meta-analyses is CAMA (community-augmented meta-analysis), a new, specialized publication format for meta-analyses. The core of such a system is the data repository, where effect sizes, completed meta-analyses, and metadata are stored and continuously curated.

The maintenance of such a repository, however, is challenging. Depending on the specific domain, a taxonomy for the concepts that are typically assessed, their designations, and the standards for the structure of the collected data has to be defined to allow the combination of research results assessing the same concepts or relations, regardless of how these were originally designated. This crucial, complex task must be undertaken for every meta-analysis to ensure that all research results are retrievable and comparable. Standards and taxonomies to ensure this is an essential aspect of a CAMA platform.

Furthermore, the continuous maintenance of a CAMA repository is both time- and labor-intensive. There are two ways to reduce the necessary workload, and these are already being applied to varying degrees in the systems presented here. The first one is crowdsourcing. MetaLab and Curate Science rely largely on this form of data accumulation. The difficulties encountered when relying on crowdsourcing, however, including how to motivate the crowd, how to educate contributors sufficiently to fulfill their tasks (e.g., by means of well-documented templates and tutorials), and how to ensure the quality of the contributions. Therefore, curation systems require quality checks, such as peer-reviewing of the added data or, as in the case of MetaLab, a curator who is responsible for checking all contributions before updating.

The second possibility to reduce the workload for the curation of the repository is the automatization of processes such as those involved in literature search (e.g., push notifications, database aggregators, automatic retrieval of full texts) and selection (e.g., machine-learning classifiers), or extraction of information from published reports (e.g., Robot Reviewer for information extraction and risk of bias assessment, Graph2Data for automatic data extraction from graphics) (Thomas et al., 2017). Currently, the

software used to carry out these tasks is far from perfect and requires manual supervision. An R package facilitating all the single tasks mentioned at once, from abstract screening to data extraction and reporting of the literature selection process, is “metagear” (Lajeunesse, 2016). However, the further development of software is a research field in its own right. Algorithms need training data to learn how to decide on the inclusion of studies and extract information from reports. These training data have to be produced by manual effort.

Thus, neither crowdsourcing nor automatization completely solves the problem of the continuous curation of cumulative, meta-analytic evidence. All relevant processes in the selection, collection, and standardization of research results require human supervision. However, this is an effort providing benefits for the research community as a whole by improving the usability and currency of existing evidence. As continuously curated meta-analytic evidence also discloses and specifies research gaps, it enables efficient distribution of research funds for closing these gaps purposefully.

References

- Baer, R., Gu, J., Cavanagh, K., & Strauss, C. (2019). Differential sensitivity of mindfulness questionnaires to change with treatment: A systematic review and meta-analysis. *Psychological Assessment*, 31(10), 1247–1263. <https://doi.org/10.1037/pas0000744>
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Editorial: evidence on questionable research practices: The good, the bad, and the ugly. *Journal of Business and Psychology*, 31(3), 323–338. <https://doi.org/10.1007/s10869-016-9456-7>
- Bastian, H., Glasziou, P., & Chalmers, I. (2010). Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? *PLoS Medicine*, 7(9), e1000326. <https://doi.org/10.1371/journal.pmed.1000326>
- Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C., & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: Insights from language acquisition research. *Child Development*, 89(6), 1996–2009. <https://doi.org/10.1111/cdev.13079>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
- Bosco, F., Steel, P., Oswald, F., Uggerslev, K., & Field, J. (2015). Cloud-based meta-analysis to bridge science and practice: Welcome to metaBUS. *Personnel Assessment and Decisions*, 1(1), 3–17. <https://doi.org/10.25035/pad.2015.002>
- Bosco, F. A., Field, J. G., Larsen, K. R., Chang, Y., & Uggerslev, K. L. (2020). Advancing meta-analysis with knowledge-management platforms: Using metaBUS in psychology. *Advances in Methods and Practices in Psychological Science*, 3(1), 124–137. <https://doi.org/10.1177/2515245919882693>
- Braver, S. L., Thoenes, F. J., & Rosenthal, R. (2014). Continuously cumulating meta-analysis and replicability. *Perspectives on Psychological Science*, 9(3), 333–342. <https://doi.org/10.1177/1745691614529796>
- Clarke, M., Brice, A., & Chalmers, I. (2014). Accumulating research: A systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. *PLoS One*, 9(7), e102670. <https://doi.org/10.1371/journal.pone.0102670>
- Cooper, H. (2017). *Research synthesis and meta-analysis. A step-by-step approach*. Sage Publications.
- Créquit, P., Trinquart, L., Yavchitz, A., & Ravaud, P. (2016). Wasted research when systematic reviews fail to provide a complete and up-to-date evidence synthesis: The example of lung cancer. *BMC Medicine*, 14(8), 1–15. <https://doi.org/10.1186/s12916-016-0555-0>
- Elliott, J. H., Synnot, A., Turner, T., Simmonds, M., Akl, E. A., McDonald, S., Salanti, G., Meerpohl, J., MacLehose, H., Hilton, J., Tovey, D., Shemilt, I., & Thomas, J. (2017). Living systematic review: 1. Introduction – the why, what, when, and how. *Journal of Clinical Epidemiology*, 91, 23–30. <https://doi.org/10.1016/j.jclinepi.2017.08.010>
- Friley, R. C., & Vazire, S. (2014). The N-pact factor: Evaluating the quality of empirical journals with respect to sample size and statistical power. *PLoS One*, 9(10), e109019. <https://doi.org/10.1371/journal.pone.0109019>
- Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Schünemann, H. J., Akl, E. A., Beyene, J., Chang, S., Churchill, R., Dearness, K., Guyatt, G., Lefebvre, C., Liles, B., Marshall, R., Martínez García, L., Mavergames, C., Nasser, M., Qaseem, A., Sampson, M., ... Schünemann, H. J. (2016). When and how to update systematic reviews: Consensus and checklist. *British Medical Journal (Online)*, 354, 1–10. <https://doi.org/10.1136/bmj.i3507>
- Haddaway, N. R. (2018). Open synthesis: On the need for evidence synthesis to embrace open science. *Environmental Evidence*, 7(1), 4–8. <https://doi.org/10.1186/s13750-018-0140-4>
- Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2019). Primary study quality in psychological meta-analyses: An empirical assessment of recent practice. *Frontiers in Psychology*, 9, Article 2667, 1–15. <https://doi.org/10.3389/fpsyg.2018.02667>
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, 19(5), 640–648. <https://doi.org/10.1097/EDE.0b013e31818131e7>
- Kepes, S., & McDaniel, M. A. (2015). The validity of conscientiousness is overestimated in the prediction of job performance. *PLoS One*, 10(10), Article e0141468. <https://doi.org/10.1371/journal.pone.0141468>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, 3(6), 643–654. <https://doi.org/10.1504/IJTEL.2011.045454>
- Lajeunesse, M. J. (2016). Facilitating systematic reviews, data extraction and meta-analysis with the metagear package for R. *Methods in Ecology and Evolution*, 7(3), 323–330. <https://doi.org/10.1111/2041-210X.12472>
- Lakens, D., Hilgard, J., & Staaks, J. (2016). On the reproducibility of meta-analyses: Six practical recommendations. *BMC Psychology*, 4(1), 1–10. <https://doi.org/10.1186/s40359-016-0126-3>
- Lakens, D., van Assen, M., Anvari, F., Corker, K., Grange, J., Gerger, H., Hasselman, F., Koyama, J., Locher, C., Miller, I., Page-Gould, E., Schönbrodt, F. D., Sharples, A., Spellman, B. A., & Zhou, S. (2017 31). *Examining the reproducibility of meta-analysis in psychology: A preliminary report*. <https://doi.org/10.31222/osf.io/xfbjf>

- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M., & Vanpeamel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402. <https://doi.org/10.1177/251524591878748>
- London, J. E. (2016). *The effect of time period, field, and coding context on rigor, interrater agreement, and interrater reliability in meta-analysis*. Dissertation, North Carolina State University.
- Millard, T., Synnot, A., Elliott, J., & Turner, T. (2018). *Results from the evaluation of the pilot living systematic reviews*. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201905_LSR_pilot_evaluation_report.pdf
- Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect–Shmozart effect: A meta-analysis. *Intelligence*, 38(3), 314–323. <https://doi.org/10.1016/j.intell.2010.03.001>
- Shojania, K. G., Sampson, M., Ansari, M. T., Ji, J., Doucette, S., & Moher, D. (2007). How quickly do systematic reviews go out of date? A survival analysis. *Annals of Internal Medicine*, 147(4), 224–233. <https://doi.org/10.7326/0003-4819-147-4-200708210-00179>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547. <https://doi.org/10.1037/a0033242>
- Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, 144(12), 1325–1346. <https://doi.org/10.1037/bul0000169>
- Synnot, A., Turner, T., & Elliott, J. (2017). *Cochrane Living Systematic Reviews*. https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/LSRInterimguidance_v0.3_20170703.pdf
- Thomas, J., Noel-Storr, A., Marshall, I., Wallace, B., McDonald, S., Mavergames, C., Glasziou, P., Shemilt, I., Synnot, A., Turner, T., & Elliott, J. (2017). Living systematic reviews: 2. Combining human and machine effort. *Journal of Clinical Epidemiology*, 91, 31–37. <https://doi.org/10.1016/j.jclinepi.2017.08.011>
- Tran, U. S., Hofer, A. A., & Voracek, M. (2014). Sex differences in general knowledge: Meta-analysis and new data on the contribution of school-related moderators among high-school students. *PLoS One*, 9(10), Article e110391. <https://doi.org/10.1371/journal.pone.0110391>
- Tsuji, S., Bergmann, C., & Cristia, A. (2014). Community-augmented meta-analyses: Toward cumulative data assessment. *Perspectives on Psychological Science*, 9(6), 661–665. <https://doi.org/10.1177/1745691614552498>
- Tsuji, S., Bergmann, C., Lewis, M., Braginsky, M., Piccinini, P., Frank, M. C., & Cristia, A. (2017). MetaLab: A repository for meta-analyses on language development, and more. In International Speech Communication Association (Ed.), *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH, 2017)*. https://www.isca-speech.org/archive/Interspeech_2017/pdfs/2053.PDF
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48. <https://doi.org/10.18637/jss.v036.i03>
- Voracek, M., Kossmeier, M., & Tran, U. S. (2019). Which data to meta-analyze, and how? A specification-curve and multiverse-analysis approach to meta-analysis. *Zeitschrift für Psychologie*, 227(1), 64–82. <https://doi.org/10.1027/2151-2604/a000357>
- Zhu, Z. N., Jiang, Y. F., & Ding, T. (2014). Risk of fracture with thiazolidinediones: An updated meta-analysis of randomized clinical trials. *Bone*, 68, 115–123. <https://doi.org/10.1016/j.bone.2014.08.010>
- Zwaan, R., Etz, A., Lucas, R., & Donnellan, M. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. <https://doi.org/10.1017/S0140525X17001972>

History

Received April 24, 2020

Revision received June 18, 2020

Accepted August 17, 2020

Published online February 4, 2021

Conflict of Interest

We have no known conflict of interest to disclose.

ORCID

Tanja Burgard

<https://orcid.org/0000-0001-9194-4821>

Michael Bošnjak

<https://orcid.org/0000-0002-1431-8461>

Tanja Burgard

Leibniz Institute for Psychology (ZPID)

Universitätsring 15

54296 Trier

Germany

burgard@leibniz-psychology.org