Taking Behavioral Science to the next Level: Opportunities for the Use of Ontologies to Enable Artificial Intelligence-Driven Evidence Synthesis and Prediction

Oscar Castro[1][a], Jacqueline Louise Mair[1,2][b], Florian von Wangenheim[1,3,*][c] and Tobias Kowatsch[4,5,3,1,*][d]

[1] *Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore*
[2] *Saw Swee Hock School of Public Health, National University of Singapore, Singapore*
[3] *Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland*
[4] *Institute for Implementation Science in Health Care, University of Zurich, Zurich, Switzerland*
[5] *School of Medicine, University of St. Gallen, St. Gallen, Switzerland*

Keywords: Taxonomy, Classification System, Ontology, Behavioral Medicine, Health Psychology, Evidence Synthesis, Systematic Review, Machine Learning.

Abstract: Decades of research have created a vast archive of information on human behavior, with relevant new studies being published daily. Despite these advances, knowledge generated by behavioral science – the social and biological sciences concerned with the study of human behavior – is not efficiently translated for those who will apply it to benefit individuals and society. The gap between what is known and the capacity to act on that knowledge continues to widen as current evidence synthesis methods struggle to process a large, ever-growing body of evidence characterized by its complexity and lack of shared terminologies. The purpose of the present position paper is twofold: (i) to highlight the pitfalls of traditional evidence synthesis methods in supporting effective knowledge translation to applied settings, and (ii) to sketch a potential alternative evidence synthesis approach which leverages on the use of ontologies – formal systems for organizing knowledge – to enable a more effective, artificial intelligence-driven accumulation and implementation of knowledge. The paper concludes with future research directions across behavioral, computer, and information sciences to help realize such innovative approach to evidence synthesis, allowing behavioral science to advance at a faster pace.

1 INTRODUCTION

Increasing physical activity, reducing greenhouse gas emissions, or avoiding antibiotic overuse. The solution to many health and environmental challenges humanity faces today lies in changing people's behavior (Ghebreyesus, 2021). Behavioral science – including a wide range of disciplines such as psychology, sociology, economics, anthropology, law, or political science – is critical in our ability to understand, predict, and ultimately shape human behavior. The success of behavioral science in reaching these aims lies in its capacity to successfully integrate and build upon evidence cumulatively. In other words, to effectively synthesize the existing body of evidence.

Evidence synthesis refers to the process of bringing together all relevant information investigating the same topic (typically scientific publications and/or datasets) to comprehensively

[a] https://orcid.org/0000-0001-5332-3557
[b] https://orcid.org/0000-0002-1466-8680
[c] https://orcid.org/0000-0003-3964-2353
[d] https://orcid.org/0000-0002-3905-2380
* These authors share senior authorship

Castro, O., Mair, J., von Wangenheim, F. and Kowatsch, T.
Taking Behavioral Science to the next Level: Opportunities for the Use of Ontologies to Enable Artificial Intelligence-Driven Evidence Synthesis and Prediction.
DOI: 10.5220/0012437300003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 2*, pages 671-678
ISBN: 978-989-758-688-0; ISSN: 2184-4305

provide an answer to a given research question (Langlois et al., 2018). Evidence synthesis often involves using a specific methodology such as systematic reviews, meta-analysis, scoping reviews, rapid reviews, umbrella reviews, qualitative reviews, or narrative reviews (hereby referred to as 'traditional evidence synthesis methods'). While the research questions addressed and the approaches used by such reviews vary considerably, they all share the principles of rigor, transparency, and accountability of methods.

Traditional evidence synthesis methods have proven to be somewhat helpful in providing trustworthy information to identify gaps in knowledge, establishing an evidence base for best-practice guidance, and helping policymakers, researchers, and the public to make more informed decisions (see, for example, the evidence-based medicine movement; Signore & Campagna, 2023). However, they also have several documented limitations (Moore et al., 2022), which hamper behavioral science development in particular and scientific progress in general. The present position paper aims to highlight these limitations and present a potential alternative evidence synthesis approach, including a research agenda moving forward.

2 KEY CHALLENGES IN USING TRADITIONAL EVIDENCE SYNTHESIS METHODS

2.1 Slow Methods to Synthesize a Rapidly Growing Body of Evidence

The amount of funds spent on research and the number of researchers globally have grown steadily over the past few decades (Lewis et al., 2021). In addition, an academic system that heavily incentivizes the publication of academic work to succeed has gained traction in many countries. As a result, the proliferation of scientific publications in the behavioral and other sciences is happening at an unparalleled rate (Bornmann et al., 2015).

With the already vast scientific literature growing exponentially, it is difficult for researchers to manually track and analyze new publications in their field. Traditional evidence synthesis methods involve a series of pre-defined steps, most of them being extremely time-consuming and labor-intensive, such as relevant study identification, data extraction, or risk of bias assessment (Borah et al., 2017). Because of the extensive work required, even in cases where solid evidence is found and synthesized, the evaluation and recommendations that follow are often unavoidably outdated by the time the review is finished (Garner et al., 2016). This does not even account for the peer-review process once the review is completed, which might take months from submission to publication. A more efficient analysis of the behavioral science literature is needed to help tackle pressing problems where human behavior plays a pivotal role.

A prime example is the COVID-19 pandemic, when governments needed to urgently understand how to encourage individuals to adhere to mask-wearing and hand-washing for the control of SARS-CoV-2 transmission. Even if a rapid review of existing evidence on individual protective behaviours was conducted, it would still take several months for a highly trained research team to produce and publish a report, by which point it would be too late to implement findings.

2.2 Prone to Human Errors and Bias

In the context of traditional evidence synthesis methods, the processes of (i) reporting study results and (ii) manually gathering such reports for evidence synthesis purposes are highly disjointed. The second step is frequently carried out by a different team than the one who conducted the study in the first place. While this facilitates impartial evidence synthesis, it also introduces a high risk of errors (Wang et al., 2020). The study report itself is already a simplification of what happened, and reviewing efforts add a second layer, further blurring what was done and found in the real world. For example, a study found a high prevalence of data extraction errors (up to 50%) in systematic reviews and, most worryingly, these often influenced effect estimates (Mathes et al., 2017). Another study analyzing the search strategy of 137 systematic reviews found that a majority (92%) incurred some type of error, such as missing key terms or wrong syntax (Salvador-Oliván et al., 2019). In addition, it is worth noting that many subjective decisions are made by researchers at different stages of a review, which increases the risk of bias (voluntary or not) and may influence various aspects of the study, such as the search strategy, selection criteria, or interpretation of findings.

2.3 Only as Good – or as Bad – as the Underlying Study Reports

Traditional evidence synthesis methods rely almost exclusively on the use of study reports. This is problematic for many reasons. First, despite the proliferation of standard reporting guidelines for different types of research (e.g., EQUATOR network; Simera et al., 2010), many reports fail to account for key aspects of the study, including population characteristics, research methods, theoretical basis, or the active components and their delivery for intervention research (Sumner et al., 2018).

Second, even if all key information about a given study is reported, authors tend to do so in a heterogeneous way. Clear labelling and classification are the basis for the organization of scientific knowledge, yet the constructs that have been established in behavioral science have not been systematized or formalized (Lawson & Robins, 2021) and often incur in 'jingle-jangle' fallacies (i.e., when one makes the erroneous assumptions that two different things are the same because they bear the same name). This phenomenon is notoriously common in behavioral science, affecting both constructs and measures (e.g., personality and values are sometimes mixed and treated as the same construct). An explanation for this might be that the terminology used in behavioral science often overlaps with informal and colloquial vocabulary, which is polysemous in nature (Hastings et al., 2022). The fact that study reporting is often inconsistent and incomplete leads to research waste (as findings cannot be integrated with other research) and could also be one of the reasons for which many meta-analyses (regarded as the gold standard method for evidence synthesis) often exhibit a high degree of heterogeneity and inconclusive results (Bryan et al., 2021). It might be possible that studies thought to be comparable were indeed different, introducing noise into the analyses (i.e., 'mixing apples and oranges').

Third, study reports leverage on text to summarize the study and communicate the results. This is logical as written language has been a main vehicle to convey information throughout human history. However, compared to numerical or other types of data, text is not universal (most publications are written in the English language), adds variability because of the different styles in writing and word-choices among researchers, and it is challenging to produce and process.

2.4 Too Many of Them

In an era of perverse academic incentives, the publication of evidence synthesis studies is also proliferating, and in some fields it even outpaces the publication of primary research (Niforatos et al., 2020). Publishing reviews has become a goal on its own, often perceived as more important than the service reviews are meant to provide. As a result, many reviews are even conducted without an adequate evidence base to justify the synthesis efforts. For example, the median number of studies included in Cochrane reviews is six to 16, and the median number of participants per trial is approximately 80 (Roberts & Ker, 2015). Systematic reviews of few trials with small sample sizes deceive the public by advertising inflated treatment effects that often become smaller or absent when more, higher-quality trials are conducted. This and other practices have led many to argue that the publication of redundant, untrustworthy, or poor-quality reviews is an increasingly unwanted contributor to the 'research waste' phenomenon (Ioannidis, 2016).

2.5 Summary

We argue the current status quo regarding evidence synthesis in behavioral science is plagued with problems caused by too much, incompletely reported evidence and mismatched conceptualizations for key constructs and measures, for which traditional evidence synthesis methods are not well-equipped nor provide an adequate and timely response. Business-as-usual is not an option if we are to build a robust scientific field to address the many 21st-century challenges for which understanding and changing human behavior is key (Hallsworth, 2023).

3 WHAT IS NEEDED TO TAKE EVIDENCE SYNTHESIS IN BEHAVIORAL SCIENCE TO THE NEXT LEVEL?

Recent advances in behavioral, information, and computing sciences could enable researchers to overcome the above-mentioned pitfalls in evidence synthesis, providing opportunities to address complex research questions efficiently and effectively. In Figure 1, we sketch a novel evidence synthesis approach that leverages ontologies to facilitate (i) complete, comprehensive, and interoperable reporting of research and (ii) a combination of human and artificial intelligence as part of the evidence synthesis process to gain speed and optimize the value of existing evidence.

An ontology is a consensually created structure for organizing knowledge in a particular domain which includes a systematic set of shared terms (controlled vocabulary) and is explicit on their interrelationships (e.g., one is part of the other). Ontologies are a result of early attempts to computationally represent and reason with knowledge and can be used to define and categorize a wide variety of constructs and terms in different fields, from wine products to Netflix content (Oliveira et al., 2021). Basically, anything that exists.

While the use of ontologies is common in the biomedical sciences and other fields (see, for example, the successful use case of the gene ontology; Gene Ontology Consortium, 2019), they have not been broadly adopted within the social and behavioral sciences (Sharp et al., 2023). More recently, the interest in the application of ontologies has been rising, including a recent consensus report by the USA National Academies of Sciences, Engineering, and Medicine on the need to develop and use ontologies to accelerate behavioral science (Sharp et al., 2023). Published ontologies in the field include the Behaviour Change Intervention Ontology (Michie et al., 2020), the Cognitive Atlas (Poldrack & Yarkoni, 2016), or the Ontology for Medically Related Social Entities (Hicks et al., 2016).

The approach depicted in Figure 1 starts with a key process consisting of coding research findings and/or datasets using ontologies (i.e., identify the presence of ontology entities as meta-data summarizing study methods and results), with the ultimate goal of creating a database – knowledge repository – which contains the evidence base of a given domain. Unlike traditional evidence synthesis methods, a key advantage of using ontologies is that entities are formally and logically connected to one another using machine-readable terms (Husáková & Bureš, 2020), and thus they enable codification of knowledge in a computer-readable format to facilitate organization, re-use, integration, and analysis (Michie et al., 2020).

A distinction is made between 'new' research findings, for which the most effective method would be to engage researchers in routinely registering their research findings using an ontology-based platform, and 'past' research findings, which would require additional efforts in case study authors are not available to register the findings retrospectively. A hybrid artificial intelligence-human information extraction tool would be needed to process these 'legacy' study reports and formalize them according to a given ontology structure (e.g., West et al., 2023).

In addition to study reports, ontologies could also be used as a framework to harmonize different datasets. Data harmonization refers to the process of reconciling various types, levels and/or sources of data in formats that are comparable and compatible. Harmonizing and integrating data from multiple sources can potentially increase the statistical power that a single dataset would provide, allowing for better decision making. There have been various efforts for cross-cohort data harmonization and integration using an ontology approach (e.g., Hao et al., 2023).

Ontologies could also help to better represent theory, which is an integral component of behavioral science. Theories so far have been communicated using natural language and thus the same issues as with study reports and the wider behavioral science literature apply (e.g., lack of clarity and consistency). Many theories overlap by referring to fundamentally the same constructs using different terms, hindering comparison and integration. Formulating theories in an ontology format provides a better basis for searching, comparing, and integrating them. It would also allow researchers to be more precise and test their propositions, as behavior science theories have been deemed to be often formulated so vaguely or abstractly that it is challenging to test or falsify them (Eronen & Bringmann, 2021). In this context, an Ontology-Based Modelling System has been already developed to formally represent 71 behavior-change theories in a way that is clear, consistent and computable (Hale et al., 2020).

Once a comprehensive database has been built, the pre-defined ontological structure would allow for the application of innovative data analysis approaches that go beyond traditional meta-analysis, including the use of artificial intelligence to answer research questions, guide future research, and inform decisions (Mac Aonghusa & Michie, 2020). For example, machine learning could be applied to predict outcomes of behavior change interventions. The system would consider subsets of previously inputted interventions based on their similarity with the scenario proposed (e.g., in terms of setting, population, intervention content, or mode of delivery) and predict an outcome value accordingly (Ganguly et al., 2022). The prediction could even take place beyond the current evidence base by allowing extrapolation of the likely outcomes of hypothetical studies (i.e., for scenarios without direct evidence).

This goes a step further compared to existing meta-analyses and meta-regressions, which focus only on estimating the average difference (e.g., mean effect size) between an intervention and comparator in an existing set of studies, and could provide a more direct answer to the questions that practitioners and policymakers typically ask (e.g., give the best estimate of the outcome for a specific scenario if we apply A or if we apply B). In addition, we are seeing a rapid advance in the development of new techniques that combine effective data-driven learning algorithms with formalizations like ontologies. Thus, it can be expected that reporting knowledge in such a way will allow for increasingly sophisticated applications able to harness the automation of learning and inference (Hastings, 2022).

Last, predictions and other outcomes resulting from this approach (Figure 1) should ideally include meta-data on the outcomes themselves, helping users understand what the level of confidence is in the outcome, as well as how much and what research has been used as part of the evidence synthesis process. Various stakeholders (e.g., health care practitioners, researchers, policymakers, educators, and students) could use such a system freely, providing feedback and helping sense check the outputs. The evidence synthesis approach described is thought to operate autonomously as much as possible, allowing researchers to shift efforts towards producing primary research or other productive tasks.

4 HOW DO WE GET THERE?

Realizing the vision represented in Figure 1 will not be a simple feat, requiring innovative, coordinated, and multidisciplinary research work on several fronts. The Human Behaviour Change Project (HBCP; 2017-2023) has been the most comprehensive effort to date assessing the feasibility of developing an artificial intelligence-based Knowledge System, including an automated information extraction component from study reports and an automated prediction component, both of which following a Behaviour Change Intervention Ontology (BCIO) developed as part of the project to provide the semantic structure for the domain (Michie et al., 2020).

The project has been instrumental in raising awareness about the need to improve evidence synthesis in behavioral science, as well as in providing first-version tools and resources that offer a step change in the possibilities for addressing evidence synthesis (e.g., an ontology, automated study identification and information extraction tools). However, the field is still in its infancy, and substantial additional efforts are needed covering various domains.

A critical first step to building the proposed evidence synthesis approach is the development of high-quality, semantically ‘strong’ ontologies (i.e., formal representation in a logic that allows specification of machine-readable properties of entities). The BCIO developed as part of the HBCP is probably the most comprehensive ontology developed in the behavioral science domain and has been highlighted in the NASEM report as a good example of a successful ontology characterized by ‘strong’ semantics. The BCIO is an ontology for all aspects of human behavior change interventions and their evaluation (e.g., where did the intervention take place?, who took part?, what Behavior Change Techniques were used?, what was the intervention schedule?), including hundreds of entities with uniquely identifiable IDs to clearly and comprehensively described what happened in a behavior change intervention. While it applies specifically to interventional research, it is a good starting point, and their development methods have been published to help inform the development of future ontologies (Wright et al., 2020). This will certainly be needed as most of the current classification systems used in behavioral science do not have formal semantics. Thus, they do not readily support automated reasoning and other artificial intelligence applications (Sharp et al., 2023).

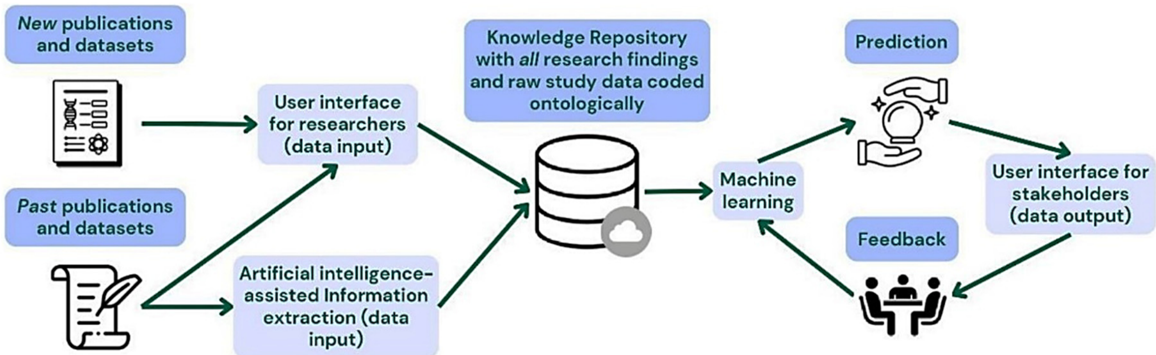

Figure 1: A hypothetical evidence synthesis approach that relies on an ontology-based Knowledge Repository system to enable the application of artificial intelligence for evidence synthesis and prediction.

Another important step is the development of tools to enable and facilitate the different processes included in the novel evidence synthesis approach, including evidence identification, ontology management and visualization, post-study registration for researchers based on an ontological structure, data infrastructure, and evidence visualization, synthesis, and prediction. The HBCP and other projects have conducted some early work in this regard. For example, study identification has progressed greatly in the past few years. It is now possible to largely automate the process of filtering new literature on a given topic (e.g., COVID-19) with high accuracy (Shemilt et al., 2021). From an initial purely manual workflow, algorithms were developed and trained to move into a position where most of the screening work is automated. As another example, the Paper Authoring Tool (PAT) is an online tool for writing up randomized controlled trial reports that prompts users for all required information (this can follow an ontological structure) and creates both Word and machine-readable JSON files upon completing the process (West, 2020). Other tools exist to semi-automate the process of assessing the risk of bias in randomized controlled trials (Jardim et al., 2022). For each tool and subprocess, researchers will need to investigate how to optimally combine human and artificial intelligence, leveraging on what each of these do best.

Specific to the computer science domain, there is also the need to investigate how to best extrapolate existing artificial intelligence methods for making predictions and synthesizing information for the database of behavioral science research. For example, a challenge to developing accurate prediction models in research is the relatively reduced number of data points for a given domain (e.g., hundreds of physical activity promotion trials) compared with the vast databases in which these models are typically trained and evaluated (e.g., billions of digitized books and newspaper articles, *all* pictures on the internet). In addition, this specific artificial intelligence application requires a certain degree of explanation for users to trust the system (e.g., confidence in the prediction, an overview of which evidence has been synthesized), which is not typically available when neural networks or other commonly employed 'black box' artificial intelligence approaches are used. The HBCP team has started developing a new method of explainable prediction that leverages a semantically constrained, rules-based system, combining aspects of symbolic and neural approaches. Early results suggest the system works with sparse data and it is not far from the predictive power of a black-box deep neural network, with the added benefit of providing substantially more transparency and explainability (Glauer et al., 2022). Regarding information extraction from 'past' research, the HBCP found a major challenge in fully automating the process of extracting comprehensive and accurate information from study reports due to the incomplete and unclear presentation of data. The advent of large language models may offer a step-change to assist humans in extracting information from study reports (Thirunavukarasu et al., 2023). Overall, we argue that behavioral science can also positively impact computer science by providing challenging use cases that require developing novel methods.

Once built, the proposed evidence synthesis approach must be thoroughly evaluated. A proof-of-concept system could be first developed and tested using a specific type of evidence synthesis (e.g., focusing on randomized controlled trials) in a particular behavior change domain (e.g., smoking cessation). Outcomes could be compared with traditional evidence synthesis methods addressing the same research questions.

Even if an effective system is developed and made available, behavior change research must investigate how the proposed evidence synthesis approach could gain broad adoption and be integrated into the standard research cycle. This might include surveying researchers to identify pain points and potential solutions, as well as exploring the ethical challenges and potential risks and liabilities of such an artificial intelligence-driven system in light of relevant regulations (e.g., the recently approved EU AI Act).

Last, a caveat of this innovative evidence synthesis approach, shared by traditional evidence synthesis methods, is that the system will be only as good as the data it operates with and is trained on (i.e., the quality of the underlying evidence). The available evidence on behavior change is not free of bias. For example, most behavioral research is conducted in high-income countries with predominantly white samples. Thus, findings might not be applicable to other contexts and ethnics groups. In addition, successful interventions are more likely to be reported and published. Current initiatives, such as preregistration of trials, could mitigate the publication bias for results that are statistically significant.

5 CONCLUSIONS

The achievement of effective, efficient, and timely synthesis of evidence in behavioral science is key to addressing the most pressing challenges of our time, from the climate emergency to the high prevalence of non-communicable diseases. Yet current evidence synthesis methods cannot keep up with a vast, rapidly growing body of evidence, which is often incomplete and ambiguous. Developing and implementing good-quality ontologies in the behavioral science domain promises to improve how evidence is organized, understood, and used. While advances have been made recently, there is still a long way to go to adapt our evidence into formats that computers can process, providing enough structured data available for large-scale learning algorithms to use. The proposed evidence synthesis approach will require close collaboration between domain experts (behavioral scientists), ontologists, and computer scientists with expertise in organizing and retrieving information. This collaboration has the potential to initiate a new phase in behavioral research in which studies are conducted, reported, and synthesized in a way that allows the easy retrieval of relevant research findings by a broad range of stakeholders, contributing to democratize human behavior knowledge.

AUTHOR CONTRIBUTIONS

Conceptualisation: OC; Writing – original draft: OC; Writing – review & editing: all authors. For definitions see http://credit.niso.org/ (CRediT).

DISCLOSURES

This project was conducted as part of the Future Health Technologies program, which was established collaboratively between ETH Zurich and the National Research Foundation, Singapore. The research is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program. FvW, TK, and JM are affiliated with the Centre for Digital Health Interventions (CDHI), a joint initiative of the Institute for Implementation Science in Health Care at the University of Zurich, the Department of Management, Technology, and Economics at ETH Zurich and the School of Medicine and Institute of Technology Management at the University of St Gallen. CDHI is funded in part by CSS, a Swiss health insurer, MTIP, a Swiss digital health investor company, and Mavie Next, an Austrian health provider. TK is also a founder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies, CSS, MTIP, and Mavie Next were not involved in this position paper. The remaining authors (OC) declare that the research was conducted without any commercial or financial relationships that could be construed as a potential conflict of interest.

REFERENCES

Borah, R., Brown, A. W., Capers, P. L., & Kaiser, K. A. (2017). Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open, 7*(2), e012545.

Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology, 66*(11), 2215-2222.

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour, 5*(8), 980-989.

Consortium, G. O. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research, 47*(D1), D330-D338.

Eronen, M. I., & Bringmann, L. F. (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science, 16*(4), 779.

Ganguly, D., Gleize, M., Hou, Y., Jochim, C., Bonin, F., Pascale, A., Johnston, M. (2021). Outcome prediction from behaviour change intervention evaluations using a combination of node and word embedding. Paper presented at the *AMIA Annual Symposium Proceedings*.

Garner, P., Hopewell, S., Chandler, J., MacLehose, H., Akl, E. A., Beyene, J., Guyatt, G. (2016). When and how to update systematic reviews: consensus and checklist. *BMJ*, 354.

Ghebreyesus, T. A. (2021). Using behavioural science for better health. *Bulletin of the World Health Organization, 99*(11), 755.

Glauer, M., West, R., Michie, S., & Hastings, J. (2022). ESC-Rules: Explainable, Semantically Constrained Rule Sets. *arXiv preprint* arXiv:2208.12523.

Hale, J., Hastings, J., West, R., Lefevre, C. E., Direito, A., Bohlen, L. C., . . . Groarke, H. (2020). An ontology-based modelling system (OBMS) for representing behaviour change theories applied to 76 theories. *Wellcome Open Research*, 5.

Hallsworth, M. (2023). A manifesto for applying behavioural science. *Nature Human Behaviour, 7*(3), 310-322.

Hao, X., Li, X., Zhang, G.-Q., Tao, C., Schulz, P. E., edu, A. s. D. N. I. i. l. u., & Cui, L. (2023). An ontology-based approach for harmonization and cross-cohort query of Alzheimer's disease data resources. *BMC Medical Informatics and Decision Making, 23*(1), 151.

Hastings, J., West, R., Michie, S., Cox, S., & Notley, C. (2022). Ontologies for the Behavioural and Social Sciences: Opportunities and challenges. *Workshop on Ontologies for the Behavioural and Social Sciences* (OntoBess 2021).

Hicks, A., Hanna, J., Welch, D., Brochhausen, M., & Hogan, W. R. (2016). The ontology of medically related social entities: recent developments. *Journal of Biomedical Semantics, 7*(1), 1-4.

Husáková, M., & Bureš, V. (2020). Formal ontologies in information systems development: a systematic review. *Information, 11*(2), 66.

Ioannidis, J. P. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta‐analyses. *The Milbank Quarterly, 94*(3), 485.

Jardim, P. S. J., Rose, C. J., Ames, H. M., Echavez, J. F. M., Van de Velde, S., & Muller, A. E. (2022). Automating risk of bias assessment in systematic reviews: a real-time mixed methods comparison of human researchers to a machine learning system. *BMC medical research methodology, 22*(1), 167.

Langlois, É. V., Daniels, K., & Akl, E. A. (2018). Evidence synthesis for health policy and systems: A methods guide. *World Health Organization*.

Lawson, K. M., & Robins, R. W. (2021). Sibling constructs: What are they, why do they matter, and how should you handle them? *Personality and Social Psychology Review, 25*(4), 344-366.

Lewis, J., Schneegans, S., & Straza, T. (2021). UNESCO Science Report: The race against time for smarter development (Vol. 2021): *Unesco Publishing*.

Mac Aonghusa, P., & Michie, S. (2020). Artificial intelligence and behavioral science through the looking glass: Challenges for real-world application. *Annals of Behavioral Medicine, 54*(12), 942-947.

Mathes, T., Klaßen, P., & Pieper, D. (2017). Frequency of data extraction errors and methods to increase data extraction quality: a methodological review. *BMC Medical Research Methodology, 17*(1), 1-8.

Michie, S., Thomas, J., Mac Aonghusa, P., West, R., Johnston, M., Kelly, M. P., O'Mara-Eves, A. (2020). The Human Behaviour-Change Project: An artificial intelligence system to answer questions about changing behaviour. *Wellcome Open Research*, 5.

Moore, R. A., Fisher, E., & Eccleston, C. (2022). Systematic reviews do not (yet) represent the 'gold standard'of evidence: A position paper. *European Journal of Pain, 26*(3), 557-566.

Niforatos, J. D., Chaitoff, A., Weaver, M., Feinstein, M. M., & Johansen, M. E. (2020). Pediatric literature shift: Growth of meta-analyses was 23 times greater than growth of randomized trials. *Journal of Clinical Epidemiology, 121*, 112-114.

Oliveira, L., Rocha Silva, R., & Bernardino, J. (2021). Wine Ontology Influence in a Recommendation System. *Big Data and Cognitive Computing, 5*(2), 16.

Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: informatics and the search for mental structure. *Annual review of psychology*, *67*, 587-612.

Roberts, I., & Ker, K. (2015). How systematic reviews cause research waste. *The Lancet, 386*(10003), 1536.

Salvador-Oliván, J. A., Marco-Cuenca, G., & Arquero-Avilés, R. (2019). Errors in search strategies used in systematic reviews and their effects on information retrieval. *JMLA, 107*(2), 210.

Sharp, C., Kaplan, R. M., & Strauman, T. J. (2023). The use of ontologies to accelerate the behavioral sciences: Promises and challenges. *Current Directions in Psychological Science, 32*(5), 418-426.

Shemilt, I., Arno, A., Thomas, J., Lorenc, T., Khouja, C., Raine, G., Wright, K. (2021). Cost-effectiveness of Microsoft Academic Graph with machine learning for automated study identification in a living map of coronavirus disease 2019 (COVID-19) research. *Wellcome Open Research*, 6, 210.

Signore, A., & Campagna, G. (2023). Evidence-based medicine: reviews and meta-analysis. *Clinical and Translational Imaging, 11*(2), 109-112.

Simera, I., Moher, D., Hirst, A., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Medicine, 8*(1), 1-6.

Sumner, J. A., Carey, R. N., Michie, S., Johnston, M., Edmondson, D., & Davidson, K. W. (2018). Using rigorous methods to advance behaviour change science. *Nature Human Behaviour, 2*(11), 797-799.

Thirunavukarasu, A. J., Ting, D. S. J., Elangovan, K., Gutierrez, L., Tan, T. F., & Ting, D. S. W. (2023). Large language models in medicine. *Nature Medicine, 29*(8), 1930-1940.

Wang, Z., Nayfeh, T., Tetzlaff, J., O'Blenis, P., & Murad, M. H. (2020). Error rates of human reviewers during abstract screening in systematic reviews. *PLoS ONE, 15*(1), e0227742.

West, R. (2020). An online Paper Authoring Tool (PAT) to improve reporting of, and synthesis of evidence from, trials in behavioral sciences. *Health Psychology, 39*(9), 846.

West, R., Bonin, F., Thomas, J., Wright, A., Mac Aonghusa, P., Gleize, M., Johnston, M. (2023). Using machine learning to extract information and predict outcomes from reports of randomised trials of smoking cessation interventions in the HBCP. *Wellcome Open Research, 8*, 452.

Wright, A. J., Norris, E., Finnerty, A. N., Marques, M. M., Johnston, M., Kelly, M. P., Michie, S. (2020). Ontologies relevant to behaviour change interventions: A method for their development. *Wellcome Open Research*, 5.