



RESEARCH ARTICLE

REVISED From smoking cessation to physical activity: Can ontology-based methods for automated evidence synthesis generalise across behaviour change domains?

[version 2; peer review: 2 approved, 1 approved with reservations]

Oscar Castro ^{1,2}, Emma Norris ³, Alison J Wright ^{1,4}, Emily Hayes¹, Ella Howes ^{1,5}, Candice Moore¹, Robert West⁶, Susan Michie ¹

¹Centre for Behaviour Change, University College London, London, England, UK
²Future Health Technologies, Singapore-ETH Centre, Campus for Research Excellence and Technological Enterprise (CREATE), Singapore, Singapore
³Health Behaviour Change Research Group, Department of Health Sciences, Brunel University London, London, England, UK
⁴Institute of Pharmaceutical Science, King's College London, London, England, UK
⁵Clinical Trials Research Unit, University of Leeds, Leeds, England, UK
⁶Department of Behavioural Science and Health, University College London, London, England, UK

v2 First published: 24 Jul 2024, 9:402
<https://doi.org/10.12688/wellcomeopenres.21664.1>
Latest published: 24 Mar 2025, 9:402
<https://doi.org/10.12688/wellcomeopenres.21664.2>

Abstract

Background

Developing behaviour change interventions able to tackle major challenges such as non-communicable diseases or climate change requires effective and efficient use of scientific evidence. The Human Behaviour-Change Project (HBCP) aims to improve evidence synthesis in behavioural science by compiling intervention reports and annotating them with an ontology to train information extraction and prediction algorithms. The HBCP used smoking cessation as the first ‘proof of concept’ domain but intends to extend its methodology to other behaviours. The aims of this paper are to (i) assess the extent to which methods developed for annotating smoking cessation intervention reports were generalisable to a corpus of physical activity evidence, and (ii) describe the steps involved in developing this second HBCP corpus.

Methods

The development of the physical activity corpus involved: (i) reviewing the suitability of smoking cessation codes already used in the HBCP,

Open Peer Review

Approval Status



	1	2	3
version 2 (revision) 24 Mar 2025		 view	 view
version 1 24 Jul 2024	 view	 view	

1. **Thomas Gültzow** , Open University of the Netherlands, Heerlen, The Netherlands
2. **John Downey** , University of Plymouth, Plymouth, UK
3. **Neil Howlett** , University of Hertfordshire, Hertfordshire, UK

Any reports and responses or comments on the article can be found at the end of the article.

(ii) defining the selection criteria and scope, (iii) identifying and screening records for inclusion, and (iv) annotating intervention reports using a code set of 200+ entities from the Behaviour Change Intervention Ontology.

Results

Stage 1 highlighted the need to modify the smoking cessation behavioural outcome codes for application to physical activity. One hundred physical activity intervention reports were reviewed, and 11 physical activity experts were consulted to inform the adapted code set. Stage 2 involved narrowing down the scope of the corpus to interventions targeting moderate-to-vigorous physical activity. In stage 3, 111 physical activity intervention reports were identified, which were then annotated in stage 4.

Conclusions

Smoking cessation annotation methods developed as part of the HBCP were mostly transferable to the physical activity domain. However, the codes applied to behavioural outcome variables required adaptations. This paper can help anyone interested in building a body of research to develop automated evidence synthesis methods in physical activity or for other behaviours.

Plain language summary

The Human Behaviour-Change Project (HBCP) wants to make it easier to gather and analyse information about how to change people's behaviour. To achieve this, the project collects reports about behaviour change interventions, code them based on an ontology (that is, a classification scheme to organise and represent information within a specific area), and use that data to train computer programs to automatically extract information from reports and make predictions. The project started with smoking cessation as a proof of concept but plans to expand to other behaviours. This study investigates how well the methods used for coding smoking cessation intervention reports could be applied to a new behaviour, physical activity, and describes how this new set of data on physical activity intervention reports was created. Building the HBCP physical activity data set involved: (1) checking if the ontology codes used for smoking cessation would work for physical activity, (2) deciding what kind of physical activity reports to include, (3) finding the reports, and (4) coding the reports using the ontology. During step 1 researchers found that some changes were needed to the codes used for smoking cessation, so they analysed 100 physical activity intervention reports and got feedback from experts to update the codes. After stages 2-4, they ended up with 111 physical activity reports coded using the ontology of behaviour change interventions. In conclusion, the methods used for coding smoking cessation intervention reports could mostly be applied to reports about physical activity

interventions, but changes were needed in relation to the target behaviour (for example, how behaviour is measured or whether researchers want people to start vs stop doing something). The aim of this report is to help others looking to build a data set to improve ways in which information on behaviour change interventions is gathered and analysed.

Keywords

ontology, taxonomy, classification system, evidence synthesis
automation, systematic review, exercise, movement behaviours



This article is included in the [Human Behaviour-Change Project \(including the APRICOT project\)](#) gateway.

Corresponding author: Oscar Castro (oscar.castro@sec.ethz.ch)

Author roles: **Castro O:** Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Norris E:** Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Wright AJ:** Conceptualization, Data Curation, Investigation, Methodology, Project Administration, Supervision, Writing – Review & Editing; **Hayes E:** Data Curation, Investigation, Writing – Review & Editing; **Howes E:** Investigation, Writing – Review & Editing; **Moore C:** Investigation, Writing – Review & Editing; **West R:** Conceptualization, Funding Acquisition, Resources, Writing – Review & Editing; **Michie S:** Conceptualization, Funding Acquisition, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: RW and SM are unpaid directors of the Unlocking Behaviour Change Community Interest Company. The rest of authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Grant information: This work was supported by Wellcome [201524; The Human Behaviour-Change Project: Building the science of behaviour change for complex intervention development]. The Human Behaviour-Change Project is funded by a Wellcome Trust collaborative award.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2025 Castro O *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Castro O, Norris E, Wright AJ *et al.* **From smoking cessation to physical activity: Can ontology-based methods for automated evidence synthesis generalise across behaviour change domains?** [version 2; peer review: 2 approved, 1 approved with reservations] Wellcome Open Research 2025, 9:402 <https://doi.org/10.12688/wellcomeopenres.21664.2>

First published: 24 Jul 2024, 9:402 <https://doi.org/10.12688/wellcomeopenres.21664.1>

REVISED Amendments from Version 1

In this updated version, we have made the following revisions to address reviewer feedback:

- Expanded the introduction to better contextualise the Human Behaviour Change Project (HBCP) approach within the broader research landscape.
- Clarified the keywords used in the literature search and specified the level in which these keywords were searched (i.e., title/abstract).
- Provided a rationale for using proprietary software to conduct the research (EPPI-Reviewer and Qualtrics).
- Included both (i) lack of pre-registration and (ii) the geographical concentration of physical activity experts consulted as study limitations and outlined plans for global engagement via the APRICOT project.
- Provided a stronger rationale for the inclusion of sedentary behaviour as part of the list of physical activity behavioural outcomes.
- Elaborated on the potential limitations of an AI-based evidence synthesis approach, as a result of the underlying evidence's quality and bias.
- Revised the title to better reflect the study's focus on evaluating ontology-based evidence synthesis across behaviour change domains.
- Amended the abstract and introduction to strengthen the study rationale, unpack the term evidence synthesis, and explain the need for an approach such as the HBCP's.
- Added further details regarding the survey for physical activity experts during stage 1, covering analysis and deployment.
- Clarified that the annotations during stage 4 resulted in iterative updates to the annotation manual as well as changes to the physical activity annotation code set.
- Revised the discussion to emphasise the study's knowledge gap and to include a wider range of literature outside the HBCP.

Any further responses from the reviewers can be found at the end of the article

Introduction

The solution to many of the health and environmental challenges that humanity faces today lies in changing people's behaviour. To achieve this, it is crucial to effectively synthesise and build upon evidence from behaviour change intervention studies cumulatively.

Evidence synthesis refers to the process of compiling relevant studies on a specific topic to answer a particular research question (Langlois *et al.*, 2018). In recent decades, researchers have placed a stronger emphasis on rigor and accountability of evidence synthesis methods – given birth to the now-standard systematic review methodology – and developed new approaches (e.g., meta-analysis) to statistically combine results from multiple studies (Thomas, 2024). These advancements have undoubtedly benefited behavioural science and other fields, for example, to help ground decisions on scientific evidence (Signore & Campagna, 2023). However, many argue that current evidence synthesis methods have inherent limitations (Moore *et al.*, 2022; Siontis & Ioannidis, 2018) and fail to leverage recent advancements in information and computing sciences which could help address research questions more effectively and efficiently (Michie & Johnston, 2017; Sharp *et al.*, 2023). Behaviour change evidence is currently extracted from study reports through a manual, lengthy, and error-prone process, without a shared conceptual and linguistic framework in the field to facilitate true cumulative knowledge. This leads to research waste (Roberts & Ker, 2015), as findings cannot be easily integrated with other research, and represents a missed opportunity to advance our understanding of behaviour change for the better.

The Human Behaviour-Change Project (HBCP) sought to address these limitations by developing an *artificial intelligence*-based Knowledge System that automatically extracts and synthesises information from intervention reports, structured by an *ontology* of *behaviour change interventions* (Michie *et al.*, 2017; Michie *et al.*, 2020a) (see Table 1 for a

Table 1. Glossary of terms used in the article.

Term	Definition	Source
Annotation	Process of coding selected parts of documents or other resources to identify the presence of ontology entities.	Michie <i>et al.</i> , 2017
Artificial intelligence	The practice of building computer programs to perform tasks that a human would reasonably regard as requiring intelligence.	Nilsson, 2014
Behaviour change intervention study	An intervention evaluation study of a behaviour change intervention scenario.	Michie <i>et al.</i> , 2020b
Behaviour change intervention scenario	A combination of attributes that are critical to understanding the intervention effects, such as the behaviour change techniques employed (intervention content), the way in which these techniques are delivered (intervention delivery), and the population and setting targeted (intervention context).	Michie <i>et al.</i> , 2020b
Entity	Anything that exists, that can be a continuant or an occurrent as defined in the Basic Formal Ontology.	Arp <i>et al.</i> , 2015
Ontology	A standardised framework providing a set of terms that can be used for the consistent annotation (or "tagging") of data and information across disciplinary and research community boundaries.	Arp <i>et al.</i> , 2015

glossary of **key terms** used in the article). The HBCP methodology involved building a corpus (i.e., compiling full texts of published intervention reports and **annotating** them according to an ontology) to provide a training and evaluation data set for the Knowledge System's information extraction and prediction algorithms. The Behaviour Change Intervention Ontology (BCIO) was developed as part of the HBCP to serve as a conceptual framework within which knowledge is structured and formally represented (Michie *et al.*, 2020b; Norris *et al.*, 2019; Wright *et al.*, 2020). The BCIO offers a set of definitions for describing key **entities** of behavioural interventions and the relationship between those entities, covering intervention content (Corker *et al.*, 2022; Marques *et al.*, 2023), engagement, population, setting (Norris *et al.*, 2020), target behaviour (Schenk *et al.*, 2024), mechanisms of action (Schenk *et al.*, 2023), and delivery, including mode of delivery (Marques *et al.*, 2020), source of delivery (Norris *et al.*, 2021), style of delivery (Wright *et al.*, 2023) and schedule. This ontology-based approach distinguishes the HBCP from other information extraction efforts in behavioural and social sciences, which typically rely on different approaches such as distant supervision (e.g., Wei *et al.*, 2022), classifiers (e.g., Zielinski & Mutschke, 2017), or neural networks (e.g., Shen *et al.*, 2022) and are mostly focused on extracting data from study abstracts (Legate *et al.*, 2024; Schmidt *et al.*, 2023), likely due to their greater availability and ease of access compared to full texts.

Smoking cessation was selected as the first 'proof of concept' domain within the HBCP because it is considered to have a higher number of high-quality trials and more homogeneous outcome measures compared to other behavioural domains (Michie *et al.*, 2017). Over 500 smoking cessation intervention evaluation reports were **annotated** using a code set of more than 200 entities from the BCIO, providing a detailed description of each smoking cessation '**intervention scenario**'. The vision for the Knowledge System is that it can automatically scan the scientific literature on smoking cessation and incorporate data from new intervention reports, using information extraction algorithms developed and trained on data from the human-annotated studies (Michie *et al.*, 2020a). Once the key information is extracted from a given intervention report into the Knowledge System, data are readily available to be exported to different evidence synthesis systems for relevant stakeholders to use. For example, see the HBCP's Outcome Prediction (<https://pred.hbcptools.org/interface/>) and Study Findings and Research Browser (<https://www.humanbehaviourchange.org/browser>) proof-of-concept tools. Prediction means that when users ask questions about scenarios of interest to them, the Knowledge System considers subsets of previously annotated entities based on their similarity with the scenario proposed (e.g., behaviour change techniques, population, etc) and predicts an outcome value accordingly (e.g., 9% quit rate at 6 months post-intervention). For further information on the HBCP's information extraction and prediction approach, see Bonin *et al.*, 2020a, Bonin *et al.*, 2020b, West *et al.*, 2023a and Hastings *et al.*, 2023.

In summary, the HBCP's ambition is to provide a fast, inexpensive and evidence-based system to extract information from study reports and provide inferences on the potential success of behaviour change interventions, facilitating accumulation and implementation of knowledge. One of the key research questions for the HBCP is to assess whether its ontology-based evidence synthesis methodology can be extended across different behavioural domains (Michie *et al.*, 2020a). To this end, a second corpus focused on physical activity intervention evaluation reports was developed as part of the project. Physical activity was selected as the next behaviour to provide a contrast with attributes already found in smoking cessation intervention evaluation reports. For example, physical activity studies tend to measure the adoption or increase of a behaviour, rather than abstinence or reduction. In addition, physical activity is thought to be a more complex domain than smoking cessation, with a wider range of outcome measures and behavioural targets.

The aims of this paper are to: (i) describe the development of a corpus of evidence relating to physical activity and (ii) evaluate the extent to which the methods developed for annotating smoking cessation intervention reports were generalisable over this second HBCP corpus. The research questions were:

- a) How well did the BCIO-based code set developed for annotating smoking cessation intervention reports work for annotating physical activity intervention reports, and
- b) What changes were needed to the code set?

The paper can also serve as a guide to help others when creating a body of evidence to automate evidence synthesis in behavioural science.

Methods

Building the HBCP physical activity corpus took place in four stages:

Stage 1: Reviewing the suitability of smoking cessation codes already used in the Human Behaviour-Change Project

Smoking cessation intervention reports were annotated according to a pre-defined, BCIO-informed code set, developed through annotations of reports and discussions with the study team, including an international expert in smoking cessation (RW). The suitability of the smoking cessation annotation code set was initially examined to assess the degree of modifications needed for application to the second behaviour of physical activity. Upon preliminary review the research team identified that smoking cessation codes relating to behavioural outcome would require substantial adaptation. These included:

Outcome (behaviour) captured the smoking behaviour defined to be targeted in a given intervention, comprising of four sub-levels: (i) *Behaviour*, specifying the overall behaviour addressed (e.g., *Tobacco use*), (ii) *Behaviour change type*,

specifying the type and direction of intended behaviour change for the smoking cessation intervention (e.g., *abstinence, reduction and quit attempt*), (iii) *Follow-up*, specifying whether smoking behaviour post-intervention was assessed as a *one-off assessment* or *repeated assessment* and the length of this follow-up, and (iv) *Behaviour assessment*, specifying the type of smoking behaviour assessment, including *subjective assessment* in self-report and informant verification and *objective assessment* in biochemical verification and observation of smoking behaviours. Last, *Outcome (behaviour) value* captured the reported values of the defined smoking outcome (behaviour) for each intervention group (e.g., % abstinent), while *Effect* captured the effect size reported in the intervention evaluation report comparing outcome behaviour values between intervention groups, including the *Effect size type* (e.g., Odds Ratio), *p value* and *95% confidence intervals*.

The adaptation of the above codes for the physical activity corpus involved the following two steps.

1.1 Identifying behavioural outcomes measured in 100 physical activity behaviour change randomised controlled trials included in Cochrane Reviews

One hundred randomised controlled trial reports of physical activity behaviour change interventions were annotated to identify the variety of behavioural outcomes they contained. All intervention reports were identified from published Cochrane reviews of physical activity (Baker *et al.*, 2015; Dobbins *et al.*, 2013; Freak-Poli *et al.*, 2013; Richards *et al.*, 2013) and sedentary behaviours (Downing *et al.*, 2018; Shrestha *et al.*, 2019), and a meta-analysis of behaviour change techniques in physical activity interventions for inactive adults (Howlett *et al.*, 2019). We included reviews on sedentary behaviour as part of the scoping work because, while sedentary behaviour is a distinct behaviour within the domain of physical activity, they form part of the same energy expenditure continuum and in many studies – particularly those using accelerometers – both are reported. Information on ‘what’, ‘when’ and ‘how’ physical activity outcomes were assessed in these intervention reports and extracted onto a standardised Excel sheet.

1.2 Seeking feedback from international experts in physical activity

Twenty-three international experts in physical activity research were invited to give feedback on the physical activity behavioural outcome codes resulting from the previous step. Experts included 14 behavioural scientists and public health stakeholders that had previously been invited to advise on the HBCP, and nine additional stakeholders identified by the project team. An online questionnaire was emailed to experts in 2018 using Qualtrics XM™ software (free alternatives include Google Forms or LimeSurvey; we chose Qualtrics as we considered it to be the superior option from a technical standpoint). The survey was designed to be completed within 20 minutes and was divided into six categories (full survey available as online supplementary material 1; West *et al.*, 2023b):

1. *Outcome (behaviour)*: Information about the type of behaviour involved (e.g., time spent engaging in moderate-to-vigorous physical activity 6 months after the start of the intervention).

2. *Behaviour change type*: The type of behaviour change targeted by the intervention (e.g., increase, decrease).
3. *Follow-up*: Information about the assessment made after either an intervention was initiated or an intervention was completed (e.g., 3 months after baseline).
4. *Behaviour assessment type*: The method by which data on the outcome behaviour is collected (e.g., self-report).
5. *Outcome (behaviour) value*: Information about the actual value for this behaviour reported in the study (e.g., mean of 2.3 hours per day).
6. *Effect*: Information about the difference between a given intervention condition and a comparator (e.g., mean difference of 20.1 minutes per day, SD of 12.6, 95% confidence interval 10.3-30.4).

Experts were asked whether they thought any codes should be changed or added within each category and, if so, which ones should be changed or added. The responses were collated by two researchers (EN & EHa), with feedback combined where applicable (i.e., merging similar responses by different experts) and discussed internally by the research team. Revisions were made to the physical activity behavioural outcomes specified in the annotation code set.

Stage 2: Defining the selection criteria and scope of the corpus

Stage 1 provided a comprehensive overview of the different physical activity behavioural outcomes used in the scientific literature and informed adaptations into the HBCP physical activity annotation code set. Stage 2 entailed discussions between the HBCP's computer science team and physical activity domain experts within the behavioural science team to narrow down the scope of the physical activity corpus. After establishing the general scope, the process of specifying the inclusion and exclusion criteria for the physical activity corpus was iterative, with the selection criteria expanded and refined as new intervention evaluation reports were reviewed.

In addition, by the time the annotation process for the physical activity corpus started, a series of technical advances implemented in EPPI-Reviewer 4 – a web-based software program used by the HBCP for managing and analysing data (Thomas *et al.*, 2020) – facilitated a more comprehensive annotation process compared to the first corpus (smoking cessation). These technical advances were discussed and informed further changes to the physical activity annotation code set. An open alternative to this software used for annotation is PDFAnno. We used EPPI-Reviewer because the research team had already expertise and training with this software and the developers were willing to change it to support our needs at no additional cost.

Stage 3: Identifying and screening physical activity behaviour change intervention reports for inclusion

3.1 Search strategy

Physical activity behaviour change intervention reports published in English were searched using Microsoft Academic Graph,

one of the biggest, most comprehensive bibliographic databases of scientific literature available at the time (Visser *et al.*, 2021) which since then has been discontinued (see OpenAlex for a suitable alternative). The search was performed on 20.01.2021 and used the following search string at the title or abstract level: MVPA or “moderate-to-vigorous physical activity” or MPA or VPA or “moderate physical activity” or “vigorous physical activity” or “strenuous physical activity” or “hard physical activity”, with reports additionally filtered using the Microsoft Academic’s built-in Randomised Controlled Trial classifier. These terms were identified through a scoping search in which the first author (OC) manually scanned 20 physical activity behaviour change intervention reports. Found reports from Microsoft Academic Graph were then exported to the reference management tool EndNote – to facilitate the processes of removing duplicates and finding full texts – and finally to EPPI-Reviewer where reports were annotated.

It is worth noting that the search process differed from a traditional systematic review. It was not the author’s intention to locate all relevant research, but to generate a somewhat random subsample of physical activity behaviour change randomised controlled trials to serve as a training set for the Knowledge System. In addition, given the broad selection criteria, it would be implausible to screen, select and annotate all the available literature. For this reason, a target corpus size was set in the first place and articles were screened for inclusion until that point. The target corpus size was based on the computer science team’s previous experience working within the smoking cessation field, which resulted in an estimation of the minimum number of intervention evaluation reports required to train the Knowledge System to extract key features. More specifically, annotating ~100 papers would theoretically allow (i) evaluating existing information extraction and prediction models (trained in smoking cessation) with a different behaviour, as well as (ii) fine-tuning the smoking cessation models on a fraction of the physical activity studies, and then testing them on the remaining studies.

3.2 Screening titles, abstracts and full texts

A total of five reviewers (OC, AW, EH_a, EH_o, CM) worked in pairs to independently screen the title & abstract of the records identified and assess whether they met the inclusion criteria. In a second step, full-text papers of retained intervention evaluation reports were examined by the same reviewers independently, with any discrepancies resolved with a consensus discussion. Disagreements that could not be resolved by consensus were discussed with the rest of the team in weekly meetings.

Stage 4: Annotating intervention attributes using a code set of 200+ entities from the Behaviour Change Intervention Ontology

4.1 Annotation process

The final step after achieving the target corpus size was to annotate the studies. Within the HBCP context, annotation refers to the process of coding selected parts of intervention reports or other resources to identify the presence of ontology entities (i.e., standardised ‘labels’ or ‘codes’ to describe relevant intervention features). For example, in the sentence

“*The mean age of participants was 21*”, the researcher would annotate the text “21” with the code “*mean age*”. This provides a machine-readable dataset which can be used to train information extraction and prediction algorithms, potentially increasing efficiency and reducing research waste in behaviour change research (Michie *et al.*, 2017).

The annotation process followed the same methodology as with the annotation of smoking cessation intervention reports: (i) developing an annotation manual iteratively and in collaboration with computer scientists, which specifies the type of data to be annotated against each code and the correct format (e.g., the amount of text to be included in the annotation), (ii) recruiting and training qualified annotators (e.g., researchers with experience in the behaviour change field), and (iii) assigning small batches of intervention evaluation reports to several pairs of annotators (OC, AW, EH_a, EH_o, CM), who annotate the reports independently and meet at the end of each batch to discuss any discrepancies between their coding. Where there were discrepancies, annotators were encouraged to consult the manual to determine the ‘correct’ way of annotating the relevant code. If the manual did not have a clear answer to the problem, this was brought to the wider team for discussion during weekly meetings (debriefing), with the manual and/or code set updated accordingly. Changes to the manual were additive (i.e., we did not change the *modus operandi* but rather expanded the manual’s instructions with some edge cases identified during the annotations). Once the coding was finalised and agreed upon, the data were included in the dataset. The HBCP physical activity annotation manual is available as an online supplementary material (File 2; West *et al.*, 2023b). Highlighted text indicates an addition to the manual as a result of the debriefings.

4.2 Physical activity annotation code set

The code set used to annotate the physical activity behaviour change intervention reports was constructed using relevant entities from the Behaviour Change Intervention Ontology (BCIO; <https://www.bcontology.org/>) and its development followed the stages described above, using the smoking cessation annotation code set as a starting point. The HBCP physical activity annotation code set is available as an online supplementary material to this paper (File 3; West *et al.*, 2023b), including a code-by-code comparison with the HBCP smoking cessation annotation code set to highlight their differences.

Results

Stage 1: Reviewing the suitability of smoking cessation codes already used in the Human Behaviour-Change Project

Modifications to the codes used to annotate smoking cessation intervention reports were discussed by the study team, with a particular focus on behavioural outcomes as these were deemed to be the most behaviour-specific codes. For example, modification types common in smoking cessation focus on decreasing behaviour in the form of abstinence or quit attempts, whereas physical activity interventions are more commonly designed to initiate, increase, or maintain activity behaviours. Types of behavioural assessment also differ between the behaviours, with device-based assessment

in the form of activity monitors (such as accelerometers and pedometers) common in physical activity but not in smoking cessation interventions. These initial ideas for modifications to the annotation code set were elaborated by reviewing 100 published physical activity behaviour change intervention reports.

1.1 Identifying behavioural outcomes measured in 100 physical activity behaviour change randomised controlled trials included in Cochrane Reviews

Physical activity behavioural outcomes described in these reports were total weekly minutes of activity (k=32) or sedentary behaviour (k=17), percentage of time spent in light physical activity or moderate-to-vigorous physical activity (k=12) or sedentary behaviour (k=7), number of physically active sessions attended in a week (k=11), number of steps (k=10), and percentage of the sample meeting the physical activity guidelines (k=6). Follow-up post-intervention was reported in 87 papers, with the majority reporting follow-up of 12 months (k=43) or 24 months (k=13). Behavioural assessment was performed by self-reported measurements (k=52), parent-report questionnaires (k=15), observation (k=3), or device-based measurements (k=48), including accelerometers (k=27), pedometers (k=15) and heart rate monitors (k=6). Note some studies reported more than one behavioural assessment and thus the sum of the above numbers do not match with the total number of studies reviewed (i.e., 100).

Considering both team discussions and the above extracted data, the annotation code set was modified as follows:

- *Behaviour type* under *Outcome (behaviour)* was modified to include *Physical activity*, with sub-levels of common intensities (Light, Moderate, Vigorous and Moderate-to-Vigorous) and Sedentary behaviour.
- *Initiation, Increase and Maintenance* of activity behaviours were added to *Behaviour change type* under *Outcome (behaviour)*.
- *Behaviour assessment type* was modified to add *Observation* (e.g., System for Observing Fitness Instruction Time (SOFIT); McKenzie *et al.*, 1992) and *Device-based assessments* including accelerometer, pedometer, inclinometer, and environmental activity sensor monitoring to capture physical activity at the area level (Roggen *et al.*, 2010).
- *Unit of measurement* was added under *Outcome (behaviour) value* to capture the specification of measurement (e.g., minutes per day, steps per week).
- Changes to *Effect* included adding *Mean Difference*, *Median Difference*, and *Cohen's d*, as these were more commonly reported in physical activity interventions.

This initial code set to annotate physical activity behavioural outcomes in intervention evaluation reports was used in the next stage.

1.2 Seeking feedback from international experts in physical activity

Of the 23 experts contacted, 11 completed the survey and were based in the UK (n=6), Australia (n=3), Canada (n=1) and South Africa (n=1). Expert responses and how these were addressed by the research team are reported as an online supplementary material (File 4; West *et al.*, 2023b). A summary of changes as a result of the expert feedback is provided below:

- *Behaviour type* under *Outcome (behaviour)* was modified to update our definition of sedentary behaviour to that of Tremblay *et al.* (2017) and to include *Walking*, as a commonly reported, specific physical activity behaviour.
- *Adherence* was added to *Behaviour change type* under *Outcome (behaviour)*.
- *Behaviour assessment type* was expanded to include *Ecological Momentary Assessment* (Liao *et al.*, 2016).
- *Observation* was moved to be a higher-level code for assessment, alongside *Subjective assessment* and *Device-based assessment*.
- *Indirect calorimetry* was added as a sub-level of *Device-based assessment*.
- *Outcome (behaviour) value* was expanded by adding *Statistical Adjustments* to capture outcome values that are weighted to improve classification of the data, such as adjustment by gender.
- *Hedges' g* was added to *Effect size type*.

Stage 2: Defining the selection criteria and scope of the corpus

The physical activity code set resulting from stage 1 was discussed with the computer science team and a decision was made to narrow down the annotations for 'outcome (behaviour)' and 'behaviour' subsections to focus on behaviour change interventions targeting moderate-to-vigorous physical activity and reporting it as a continuous variable. This was because incorporating different physical activity outcomes would have resulted in a higher number of intervention evaluation reports and annotations being required for training the Knowledge System to recognise and extract such outcomes, greatly increasing the required corpus size.

Moderate-to-vigorous physical activity was prioritised as it has been the main focus of physical activity and public health efforts during the past decades and has the strongest links with both physical and psychological outcomes, compared to other forms of physical activity such as light intensity physical activity or sedentary behaviour (Owen *et al.*, 2020). In addition, similar to the smoking cessation corpus, we decided to focus on randomised controlled trials due to their recognition as 'gold-standard' for studying intervention effectiveness (Michie *et al.*, 2017). A complete overview of the selection criteria for the physical activity corpus is available in Table 2.

Table 2. Selection criteria for the intervention reports included in the HBCP physical activity corpus.

Inclusion criteria	Exclusion criteria
Population <ul style="list-style-type: none"> Any age groups. Healthy individuals as well as people with physical/mental health conditions. 	n/a
Research design <ul style="list-style-type: none"> Randomised controlled trials (including pilot RCTs). 	Research design <ul style="list-style-type: none"> Quasi-experimental trials, protocols, qualitative research and economic or process evaluations.
Study aim <ul style="list-style-type: none"> Behaviour change interventions targeting physical activity. 	Study aim <ul style="list-style-type: none"> Epidemiology studies, secondary analyses, analysis of physical activity correlates.
Outcome <ul style="list-style-type: none"> Total moderate-to-vigorous physical activity (MVPA), reported as units of time. Assessed through self-report and/or device-based measures. 	Outcome <ul style="list-style-type: none"> Studies focused on steps, total physical activity, light-intensity physical activity or sedentary behaviour. Studies focused on moderate physical activity only, or that report moderate and vigorous physical activity separately. Studies focused on specific periods of the day (e.g., MVPA during PE classes only) or specific types of MVPA (e.g., household MVPA, leisure MVPA, transport MVPA). Studies where MVPA is reported as change scores (i.e., no pre- and post-test values available, just change values from baseline).
n/a	Other <ul style="list-style-type: none"> Conference submissions, PhD thesis, pre-prints and/or abstract-only entries. Studies published in languages other than English. Studies with more than 8 arms.* Study reports with physical activity results only available in figures/graphs (i.e., where no numerical data can be extracted / annotated) or rotated tables.*

*The rationale for these selection criteria reflects limitations of the software used to annotate intervention reports (EPPI-Reviewer).

In relation to the technical advances in EPPI-Reviewer by the time the annotation process for the physical activity corpus started, these included:

- The possibility to annotate outcome measures at different time points (i.e., pre, post and follow-up measurements).
- A new way to capture outcome values, incorporating the outcome values in a table, as well as their standard deviation and the number of participants per group. Where this data was available, the EPPI-Reviewer software automatically calculated the effect size(s) for the difference(s) between groups). Because effect size was now automatically calculated, it was no longer manually annotated.

These technical advances were incorporated into the physical activity annotation manual and code set.

Stage 3: Identifying and screening physical activity behaviour change intervention reports for inclusion.

A minimum corpus size of ~100 intervention evaluation reports was first established, with batches of articles reviewed

for inclusion up to achieving the target (Figure 1). Because articles were reviewed in batches, the ultimate included sample size was 111 reports.

Stage 4: Annotating intervention attributes using a code set of 200+ entities from the Behaviour Change Intervention Ontology.

A total of 111 physical activity behaviour change intervention reports published between 2005 and 2020 were annotated (see included reports in [online supplementary material 5](#); [West et al., 2023b](#)). Annotations for each of the 111 study reports (JSON file) can be found in the online supplementary materials ([File 6](#); [West et al., 2023b](#)). In addition, the HBCP has developed a Research Browser Tool (<https://www.humanbehaviourchange.org/browser>) which allows users to visualise the annotations and use the BCIO to structure queries and locate relevant studies.

The code set was slightly modified as a result of the annotation work and iterative discussions with the research team. This included:

- Office facility* was added under *Setting* to capture work-based physical activity interventions.

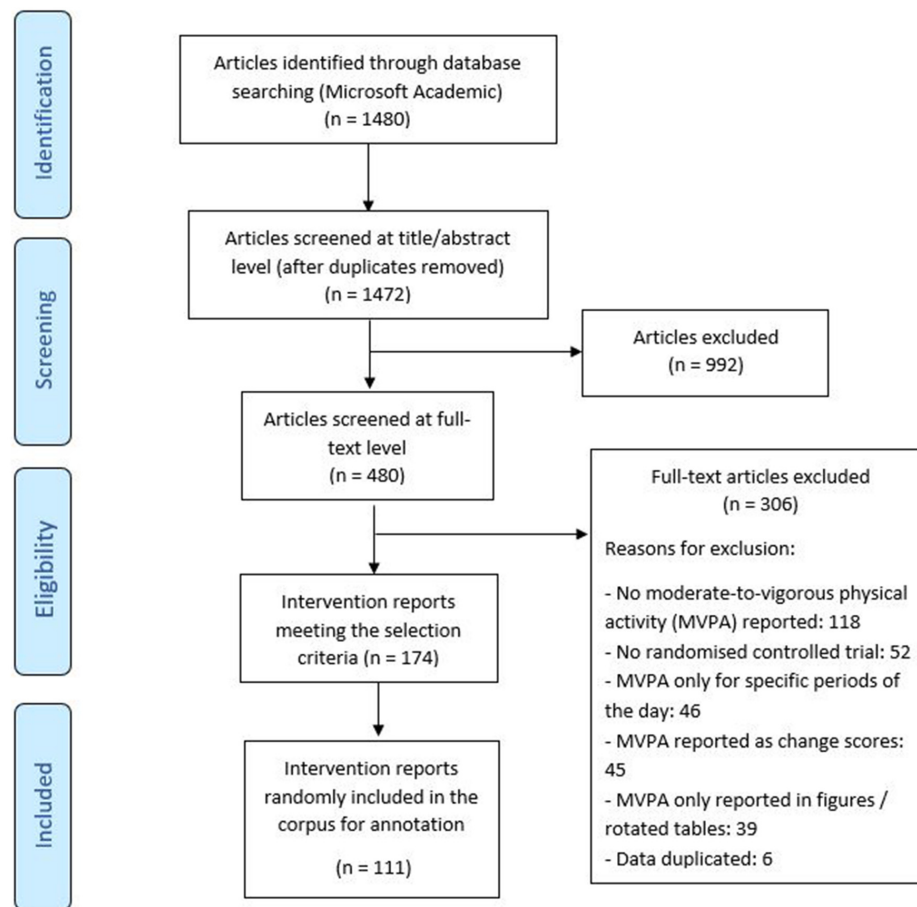


Figure 1. Flow diagram for the intervention reports included in the HBCP physical activity corpus.

- *Aggregate body mass index (BMI)* was added under *Population* as BMI is particularly relevant for physical activity (e.g., people with high BMI find unique barriers to physical activity) and thus is typically reported in physical activity behaviour change interventions.
- *Funding* and *Competing interests* were expanded to include *Industry with financial interest in intervention success* to capture when physical activity behaviour change intervention studies are supported by makers of devices used in interventions (e.g., Fitbit).

Discussion

The present paper outlined the steps taken to develop the HBCP corpus of physical activity behaviour change interventions. The main rationale behind developing this corpus was to expand on previous HBCP work and assess the extent to which the methods developed for annotating smoking cessation intervention reports were generalisable over a corpus of evidence relating to a new, and arguably more ‘challenging’ behaviour: physical activity. This is critical to investigate whether the

HBCP’s methodology can be applied to improve evidence synthesis across different behavioural domains.

Our process of adapting the annotation code set found that a majority of codes were reusable and can be effectively used to annotate physical activity interventions and potentially other behaviours (e.g., those relating to population, intervention content or mode of delivery). However, findings also underscored certain codes which would require a domain-specific approach, particularly those related to behavioural outcomes such as assessment or modification type. This is consistent with previous literature highlighting the wide variety of outcomes which are relevant to human behaviour (Larsen *et al.*, 2021; Mallpress, 2022) and suggests that a high degree of granularity is potentially needed within the BCIO to be able to capture these intervention attributes. The currently published BCIO is intended to provide a stable organising structure within which new entities can be added via an open-access portal (<https://www.bciontology.org/contribute>), allowing for tracking and “versioning” as the ontology is revised and updated as required.

By creating a second corpus we also hoped to examine whether the information extraction and prediction algorithms developed with smoking cessation studies could be applied to other behaviours. This “transfer learning” – the ability to leverage knowledge gained from one task and apply it to another, often using data from different sources (Peng *et al.*, 2024) – could enhance the performance of prediction algorithms by integrating evidence from different domains. Additionally, it could reduce the need for extensive human annotations to train information extraction algorithms by reusing pre-existing models, ultimately improving the portability of the HBCP methodology. However, due to variation and ambiguity in the way information is presented in study reports, the information extraction algorithms developed for smoking cessation had limited performance for automated information extraction and for associating information with individual study arms, which posed an insurmountable barrier to full automation (West *et al.*, 2023a). Therefore, the planned comparisons between the smoking cessation and physical activity domains in terms of accuracy of information extraction and prediction algorithms were not performed. The annotated HBCP corpora developed for smoking cessation and physical activity could still be used in the future by other research teams pursuing a similar approach to automated evidence synthesis.

Recommendations for future research

Our approach to creating the HBCP physical activity corpus and adapting the annotation code set to a new behaviour change domain can serve as a guide for those interested in building a corpus of intervention reports for automating evidence synthesis in behavioural science. We include below some recommendations for future research.

First, the groundwork conducted as part of the corpus development process (stage 1) proved crucial to systematically identify and classify relevant outcomes, assess the breadth of the field, and help take pragmatic decisions on the corpus’ scope. We recommend undertaking a thorough and systematic outcome identification process for any given behaviour before working towards information extraction automation. Related to this point, a challenge we encountered with our physical activity corpus was the wide variety of different outcomes within the physical activity field. Physical activity is often reported using different variations of physical activity intensities (light, moderate, vigorous or moderate-to-vigorous), but also as number of steps or metabolic equivalents (METs), all using different time frames and frequency metrics such as minutes per day, hours per week, or number of exercise sessions per week (Sylvia *et al.*, 2014). This makes evidence synthesis in general, and automated artificial intelligence-based evidence synthesis in particular, more difficult. The more heterogeneous a field is, the more data are theoretically required to train an artificial intelligence system working with such behaviour, due to the necessity of having enough examples to ‘teach’ the system how to recognise and extract a given type of outcome entity (e.g., minutes of weekly moderate-to-vigorous physical activity as opposed to number of steps per day). This led us to prioritise a single outcome of interest for our corpus (i.e., moderate-to-vigorous physical activity) and should be considered by future research teams attempting a similar approach for evidence synthesis automation.

An element that may ease future automation attempts in heterogeneous domains, such as physical activity, is to design the Knowledge System in a way that is able to recognise different types of physical activity and perform transformations to the extracted outcomes via pre-specified arithmetic operations. For example, the Knowledge System could be programmed to automatically sum up minutes spent in moderate and vigorous physical activity into a single variable (i.e., moderate-to-vigorous physical activity), improving inter-study operability. Another example would be for the Knowledge System to be able to harmonise the outcomes that are reported in different time frames (e.g., automatically transform hours of moderate-to-vigorous physical activity into minutes or vice versa). The HBCP’s Knowledge System was not developed in a way which allowed these operations.

Regardless of the heterogeneity in outcomes, creating a body of evidence to train information extraction and prediction algorithms will always require finding relevant intervention evaluation reports in the first place. In this regard, we highlight the use of large-scale data sets of scholarly publications (e.g., OpenAlex) as a useful tool to locate intervention reports. Compared to traditional database searching, this approach enables researchers to access a wide range of databases in the same platform, saving time and facilitating automated study identification and incorporation into the Knowledge System, which is key to ensure the system is constantly up to date.

Once the corpus of intervention evaluation reports has been created, it is important to consider that the annotation process takes a substantial amount of time and human resources. Although automated methods hold promise to improve the efficiency of data synthesis over the long term, initial human labour is required to develop and train accurate information extraction algorithms. While annotating intervention reports in pairs is important to ensure high-quality training data, one option if researchers have limited time and resources is to move to single coding once interrater reliability is acceptable and there is a complete, well-developed annotation guide.

Although good annotation tools and processes are important, the production of high-quality training data relies heavily on the data available for annotation. Behaviour change intervention reports, however, tend to use unclear and ambiguous language and this often makes it difficult to accurately interpret and classify data (Castro *et al.*, 2024; West *et al.*, 2023a). The HBCP found that intervention reports need to be much more structured and consistent in the way they present data. New authoring tools, such as the Paper Authoring Tool (PAT; West, 2020), can be implemented to produce consistent, complete and computer-readable reporting of trials, contributing to improve the extraction and synthesis of data from study reports.

Last, it is also important to recognise that artificial intelligence systems are only as good as the data they are trained on and operate with. The available evidence on behaviour change is not free of bias. For example, successful interventions are more likely to be reported and published (Nissen *et al.*, 2016). Similarly, most behavioural research is conducted in

high-income countries (Miranda & Zaman, 2010), with predominantly white samples (Oh *et al.*, 2015). Thus, findings might not be applicable to other contexts and ethnic groups.

Strengths and limitations

A strength of this research is the systematic, multi-stage process followed to develop the physical activity annotation code set and corpus (including review of intervention reports and international expert consultation) to test the generalisation of annotation methods developed for smoking cessation. Moreover, two researchers independently carried out the screening and annotation of intervention reports, reducing the risk of human error and maximising reliability.

A limitation is the fact that only intervention evaluation reports published in English were considered for inclusion. This means that the annotation methods described here, and the potential information extraction algorithms resulting from such methods, could only be employed with reports in the English language. It is also worth acknowledging that the physical activity corpus size is smaller and has a narrower scope compared to the smoking cessation corpus, meaning findings may not be applicable to all physical activity intervention research. Our intention, however, was not to create a stand-alone corpus but to build upon existing methods applied to smoking cessation intervention reports and use the new domain as a testing ground. Another limitation of this study is the geographical concentration of physical activity experts consulted, primarily from English-speaking countries and the Global North. This focus may limit the generalisability of the findings, as perspectives and contexts from other regions may not be fully represented. As part of the upcoming Advancing Prevention Research In Cancer through Ontology Tools (APRICOT) project – which extends the HBCP work by developing tools for using ontologies in behavioural science (Michie *et al.*, 2024) – we plan to set up a diverse Community of Practice to ensure broader global engagement in future work.

Last, while this study was not pre-registered due to its largely descriptive nature – lacking hypotheses or analyses susceptible to p-hacking or other questionable research practices – we acknowledge that pre-registration could still have been beneficial (e.g., to enhance discoverability and help prevent unintentional duplication of research efforts). Future hypothesis-driven research conducted using this published corpus, such as training and evaluating information extraction algorithms, should be pre-registered to ensure transparency and minimise publication bias.

Conclusions

It is possible to generalise the HBCP methods developed for annotating smoking cessation intervention reports to physical activity and potentially other behavioural domains, provided domain-specific groundwork is previously conducted (particularly in relation to behavioural outcomes). This paper provides a blueprint for anyone interested in building a body of research to enhance evidence synthesis in the physical activity

field and beyond, including (i) an ontology-informed code set for annotating physical activity behaviour change interventions, (ii) an openly available corpus of 111 annotated physical activity behaviour change interventions which could be used to train and evaluate information extraction algorithms, and (iii) recommendations for future automated evidence synthesis efforts moving forward.

Ethics

Ethical approval was granted by University College London's ethics committee (CEHP/2020/579) in February 2020. Participant consent (consultation with physical activity experts) was provided in a dedicated page of the online Qualtrics survey.

Data availability

Underlying data

Open Science Framework: Human Behaviour-Change Project. <https://doi.org/10.17605/OSF.IO/EFP4X> (West *et al.*, 2023b)

Extended data

Online supplementary materials cited in this article are available below:

- Supplementary material 1: Questionnaire for physical activity experts (<https://osf.io/9vwye/>).
- Supplementary material 2: HBCP physical activity annotation manual (<https://osf.io/8ekfz>).
- Supplementary material 3: Comparison between HBCP physical activity and smoking cessation annotation code sets (<https://osf.io/n3e9y>).
- Supplementary material 4: Responses to physical activity experts feedback (<https://osf.io/n56kj/>).
- Supplementary material 5: Intervention reports included in HBCP physical activity corpus (<https://osf.io/kdmwe>).
- Supplementary material 6: Annotations for 111 intervention reports included in HBCP physical activity corpus (<https://osf.io/dtn6u>).

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original data is properly cited.

Acknowledgements

We thank Catalina Trama Alvarez and Zoe Zhang for coding physical activity outcomes in intervention reports. We are also grateful to Elizabeth Corker for helping with the physical activity annotations and to James Thomas and Clement Veall for their assistance in identifying and screening the intervention reports. Finally, we thank Catherine Draper, Corneel Vandelanotte, Karen Milton, Mark Beauchamp, Nanette Mutrie, Neville Owen, Nick Cavill, Paul Kelly, Russ Jago, Stuart Biddle and Tessa Strain for providing expert stakeholder feedback.

References

- Arp R, Smith B, Spear AD: **Building ontologies with basic formal ontology**. Massachusetts: MIT Press, 2015.
[Reference Source](#)
- Baker PRA, Francis DP, Soares J, *et al.*: **Community wide interventions for increasing physical activity**. *Cochrane Database Syst Rev*. 2015; **1**(1): CD008366.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bonin F, Gleize M, Finnerty A, *et al.*: **HBCP corpus: a new resource for the analysis of Behavioural Change Intervention reports**. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. 2020b; 1967–1975.
[Reference Source](#)
- Bonin F, Gleize M, Hou Y, *et al.*: **Knowledge extraction and prediction from behavior science Randomized Controlled Trials: a case study in smoking cessation**. *AMIA Annu Symp Proc*. American Medical Informatics Association, 2020a; 2020: 253–262.
[PubMed Abstract](#) | [Free Full Text](#)
- Castro O, Mair JL, von Wangenheim F, *et al.*: **Taking behavioral science to the next level: opportunities for the use of ontologies to enable artificial intelligence-driven evidence synthesis and prediction**. In: *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies*. 2024; **2**: 671–678.
[Publisher Full Text](#)
- Corker E, Marques MM, Johnston M, *et al.*: **Behaviour change techniques taxonomy v1: feedback to inform the development of an ontology [version 2; peer review: 2 approved]**. *Wellcome Open Res*. 2022; **7**: 211.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dobbins M, Husson H, DeCorby K, *et al.*: **School-based physical activity programs for promoting physical activity and fitness in children and adolescents aged 6 to 18**. *Cochrane Database Syst Rev*. 2013.
[Publisher Full Text](#)
- Downing KL, Hnatiuk JA, Hinkley T, *et al.*: **Interventions to reduce sedentary behaviour in 0–5 year olds: a systematic review and meta-analysis of randomised controlled trials**. *Br J Sports Med*. 2018; **52**(5): 314–321.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Freak-Poli RLA, Cumpston M, Peeters A, *et al.*: **Workplace pedometer interventions for increasing physical activity**. *Cochrane Database Syst Rev*. 2013; (4): CD009209.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Hastings J, Glauer M, West R, *et al.*: **Predicting outcomes of smoking cessation interventions in novel scenarios using ontology-informed, interpretable machine learning [version 1; peer review: 1 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2023; **8**: 503.
[Publisher Full Text](#)
- Howlett N, Trivedi D, Troop NA, *et al.*: **Are physical activity interventions for healthy inactive adults effective in promoting behavior changes and maintenance, and which behavior change techniques are effective? a systematic review and meta-analysis?** *Transl Behav Med*. 2019; **9**(1): 147–157.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Langlois EV, Daniels K, Akl EA: **Evidence synthesis for health policy and systems: a methods guide**. Geneva: World Health Organization, 2018.
[Reference Source](#)
- Larsen KR, Ramsay LJ, Godinho CA, *et al.*: **IC-Behavior: an interdisciplinary taxonomy of behaviors**. *PLoS One*. 2021; **16**(9): e0252003.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Legate A, Nimón K, Noblin A: **(Semi) automated approaches to data extraction for systematic reviews and meta-analyses in social sciences: a living review [version 1; peer review: 2 approved, 1 approved with reservations]**. *F1000Res*. 2024; **13**: 664.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Liao Y, Skelton K, Dunton G, *et al.*: **A systematic review of methods and procedures used in ecological momentary assessments of diet and physical activity research in youth: an adapted STROBE checklist for reporting EMA studies (CREAMS)**. *J Med Internet Res*. 2016; **18**(6): e151.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Mallpress DE: **Towards a functional classification of behaviour: a taxonomy based on outcomes**. *Adapt Behav*. 2022; **30**(5): 417–450.
[Publisher Full Text](#)
- Marques MM, Carey RN, Norris E, *et al.*: **Delivering Behaviour Change Interventions: development of a mode of delivery ontology [version 2; peer review: 2 approved]**. *Wellcome Open Res*. 2020; **5**: 125.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marques MM, Wright AJ, Corker E, *et al.*: **The behaviour change technique ontology: transforming the behaviour change technique taxonomy v1 [version 1; peer review: 4 approved]**. *Wellcome Open Res*. 2023; **8**: 308.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- McKenzie TL, James FS, Nader PR: **SOFIT: System for Observing Fitness Instruction Time**. *J Teach Phys Educ*. 1992; **11**(2): 195–205.
[Publisher Full Text](#)
- Michie S, Johnston M: **Optimising the value of the evidence generated in implementation science: the use of ontologies to address the challenges**. *Implement Sci*. 2017; **12**(1): 131.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Michie S, Thomas J, Johnston M, *et al.*: **The Human Behaviour-Change Project: harnessing the power of artificial intelligence and machine learning for evidence synthesis and interpretation**. *Implement Sci*. 2017; **12**(1): 121.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Michie S, Thomas J, Mac Aonghusa P, *et al.*: **The Human Behaviour-Change Project: an artificial intelligence system to answer questions about changing behaviour [version 1; peer review: not peer reviewed]**. *Wellcome Open Res*. 2020a; **5**: 122.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Michie S, West R, Finnerty AN, *et al.*: **Representation of Behaviour Change Interventions and their evaluation: development of the upper level of the Behaviour Change Intervention Ontology [version 1; peer review: 1 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2020b; **5**: 123.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Michie S, West R, Hastings J, *et al.*: **The Human Behaviour-Change Project phase 2: advancing behavioural and social sciences through ontology tools [version 1; peer review: not peer reviewed]**. *Wellcome Open Res*. 2024; **9**: 730.
[Publisher Full Text](#)
- Miranda JJ, Zaman MJ: **Exporting “failure”: why research from rich countries may not benefit the developing world**. *Rev Saude Publica*. 2010; **44**(1): 185–189.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Moore RA, Fisher E, Eccleston C: **Systematic reviews do not (yet) represent the ‘gold standard’ of evidence: a position paper**. *Eur J Pain*. 2022; **26**(3): 557–566.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nilsson NJ: **Principles of Artificial Intelligence**. USA, Massachusetts: Morgan Kaufmann Publishers, 2014.
[Reference Source](#)
- Nissen SB, Magidson T, Gross K, *et al.*: **Publication bias and the canonization of false facts**. *Elife*. 2016; **5**: e21451.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Norris E, Finnerty AN, Hastings J, *et al.*: **A scoping review of ontologies related to Human Behaviour Change**. *Nat Hum Behav*. 2019; **3**(2): 164–172.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Norris E, Marques MM, Finnerty AN, *et al.*: **Development of an Intervention setting ontology for behaviour change: specifying where interventions take place [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2020; **5**: 124.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Norris E, Wright AJ, Hastings J, *et al.*: **Specifying who delivers Behaviour Change Interventions: development of an Intervention source ontology [version 1; peer review: 2 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2021; **6**: 77.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oh SS, Galanter J, Thakur N, *et al.*: **Diversity in clinical and biomedical research: a promise yet to be fulfilled**. *PLoS Med*. 2015; **12**(12): e1001918.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Owen N, Healy GN, Dempsey PC, *et al.*: **Sedentary behavior and public health: integrating the evidence and identifying potential solutions**. *Annu Rev Public Health*. 2020; **41**: 265–287.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Peng C, Yu Z, Smith KE, *et al.*: **Improving generalizability of extracting Social Determinants of Health using Large Language Models through prompt-tuning**. *arXiv preprint arXiv: 2403.12374*. 2024.
[Publisher Full Text](#)
- Richards J, Hillsdon M, Thorogood M, *et al.*: **Face-to-face interventions for promoting physical activity**. *Cochrane Database Syst Rev*. 2013; (9): CD010392.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Roberts I, Ker K: **How systematic reviews cause research waste**. *Lancet*. 2015; **386**(10003): 1536.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Roggen D, Calatroni A, Rossi M, *et al.*: **Collecting complex activity datasets in highly rich networked sensor environments**. *Seventh International Conference on Networked Sensing Systems*. 2010.
[Publisher Full Text](#)
- Schenk PM, West R, Castro O, *et al.*: **An ontological framework for organising and describing behaviours: the human behaviour ontology [version 1; peer review: 1 approved with reservations]**. *Wellcome Open Res*. 2024; **9**: 237.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schenk PM, Wright AJ, West R, *et al.*: **An ontology of mechanisms of action in Behaviour Change Interventions [version 1; peer review: 2 approved]**. *Wellcome Open Res*. 2023; **8**: 337.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Schmidt L, Finnerty Mutlu AN, Elmore R, *et al.*: **Data extraction methods for systematic review (semi)automation: update of a living systematic review [version 2; peer review: 3 approved]**. *F1000Res*. 2023; **10**: 401.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Sharp C, Kaplan RM, Strauman TJ: **The use of ontologies to accelerate the behavioral sciences: promises and challenges**. *Curr Dir Psychol Sci*. 2023; **32**(5): 418–426.

[Publisher Full Text](#)

Shen S, Jiang C, Hu H, *et al.*: **A model for the identification of the functional structures of unstructured abstracts in the social sciences**. *Electron Libr*. 2022; **40**(6): 680–697.

[Publisher Full Text](#)

Shrestha N, Grgic J, Wiesner G, *et al.*: **Effectiveness of interventions for reducing non-occupational sedentary behaviour in adults and older adults: a systematic review and meta-analysis**. *Br J Sports Med*. 2019; **53**(19): 1206–1213.

[PubMed Abstract](#) | [Publisher Full Text](#)

Signore A, Campagna G: **Evidence-based medicine: reviews and meta-analysis**. *Clin Transl Imaging*. 2023; **11**(2): 109–112.

[Publisher Full Text](#)

Siontis KC, Ioannidis JPA: **Replication, duplication, and waste in a quarter million systematic reviews and meta-analyses**. *Circ Cardiovasc Qual Outcomes*. 2018; **11**(12): e005212.

[PubMed Abstract](#) | [Publisher Full Text](#)

Sylvia LG, Bernstein EE, Hubbard JL, *et al.*: **A practical guide to measuring physical activity**. *J Acad Nutr Diet*. 2014; **114**(2): 199–208.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Thomas J: **Methods development in evidence synthesis: a dialogue between science and society**. Edward Elgar Publishing, 2024; 146–158.

[Publisher Full Text](#)

Thomas J, Graziosi S, Brunton J, *et al.*: **EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis**. EPPI-Centre Software. London: UCL Social Research Institute, 2020.

[Reference Source](#)

Tremblay MS, Aubert S, Barnes JD, *et al.*: **Sedentary Behavior Research Network (SBRN) -terminology consensus project process and outcome**. *Int J*

Behav Nutr Phys Act. 2017; **14**(1): 75.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Visser M, van Eck NJ, Waltman L: **Large-scale comparison of bibliographic data sources: scopus, web of science, dimensions, crossref, and microsoft academic**. *Quant Sci Stud*. 2021; **2**(1): 20–41.

[Publisher Full Text](#)

Wei X, Salsabil L, Wu J: **Theory entity extraction for social and behavioral sciences papers using distant supervision**. In: *Proceedings of the 22nd ACM Symposium on Document Engineering*. September, 2022; 1–4.

[Publisher Full Text](#)

West R: **An online Paper Authoring Tool (PAT) to improve reporting of, and synthesis of evidence from, trials in behavioral sciences**. *Health Psychol*. 2020; **39**(9): 846–850.

[PubMed Abstract](#) | [Publisher Full Text](#)

West R, Bonin F, Thomas J, *et al.*: **Using machine learning to extract information and predict outcomes from reports of randomised trials of smoking cessation interventions in the Human Behaviour-Change Project [version 1; peer review: 2 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2023a; **8**: 452.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

West R, Michie S, Shawe-Taylor J: **Human Behaviour-Change Project**. [Dataset], 2023b.

<http://www.doi.org/10.17605/OSF.IO/EFP4X>

Wright AJ, Norris E, Finnerty AN, *et al.*: **Ontologies relevant to Behaviour Change Interventions: a method for their development [version 3; peer review: 2 approved, 1 approved with reservations]**. *Wellcome Open Res*. 2020; **5**: 126.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Wright AJ, Zhang L, Howes E, *et al.*: **Specifying how intervention content is communicated: development of a Style of Delivery Ontology [version 1; peer review: 1 approved]**. *Wellcome Open Res*. 2023; **8**: 456.

[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Zielinski A, Mutschke P: **Mining social science publications for survey variables**. In: *Proceedings of the second workshop on NLP and computational social science*. August, 2017; 47–52.

[Publisher Full Text](#)

Open Peer Review

Current Peer Review Status: ? ✓ ✓

Version 2

Reviewer Report 11 April 2025

<https://doi.org/10.21956/wellcomeopenres.26422.r121859>

© 2025 Howlett N. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Neil Howlett 

Psychology, Sports and geography, University of Hertfordshire, Hertfordshire, UK

This article clearly outlines an approach to build a physical activity corpus with the potential to be helpful to other researchers trying to apply this method, and for future evidence synthesis. The article will be useful to readers and breaks down the process with an appropriate level of detail, which is still accessible enough to those not as familiar with ontologies and the HBCP in general. I note that this is version 2, which has already been revised based on the feedback from two reviewers. I have some minor suggestions, which I hope will add some additional clarity and context.

Detailed comments:

In the second sentence of the plain language summary, the words 'code' and 'use' should be plural.

At the end of the second paragraph the current process of data extraction is described as 'error-prone'. It is clear why it is described as 'lengthy' and 'without a shared conceptual and linguistic framework', but it is less clear why it is error prone. Does this refer to basic data extraction mistakes from reviewers, inconsistent coding of the same content across review teams (i.e., BCTs), or some other facet? Some context, and ideally a reference to support this assertion, would be useful here. Some commentary, either here or in the discussion, about why the proposed approach will reduce the prevalence of these errors, would be useful too. As already stated in the discussion, intervention reports tend to use unclear and ambiguous language and this often makes it difficult to accurately interpret and classify data. Are there other elements that introduce errors that are specific to manual data extraction, particularly when extraction by two reviewers takes place?

I note that in version 2 of this manuscript, the authors added justification to stage 1 for including reviews on sedentary behaviour interventions. This is useful context and using an inclusive approach at the stage of screening reviews for behavioural outcomes makes sense. However, I would argue that this is not the same domain as physical activity for three reasons:

- Much of the rationale for exploring the physical activity domain was that this behaviour involves the initiation, increase, or maintenance of behaviour (as opposed to smoking where a reduction or cessation is targeted). Sedentary behaviour interventions target a reduction in the behaviour and, therefore, it could be argued that sedentary behaviour is closer to the smoking domain in terms of the behaviour change type being targeted.
- There is a growing body of research that shows that sedentary behaviour can affect health independently of physical activity levels. For example, there are many people who meet the physical activity guidelines, but that are also highly sedentary (i.e., the rest of their week is spent sitting in front of a laptop at work or television in leisure time).
- Although there are some shared behavioural outcomes between physical activity and sedentary behaviour, they also have ways of measuring these outcomes that are unique (e.g., change in METs per day or week is not used for sedentary behaviour).

I am not suggesting that the method or results should be changed, but there should be an acknowledgement, probably in the discussion, that this approach probably needs to be done again in the future for sedentary behaviour as its own domain distinct from physical activity.

At the bottom of page 8, it is stated that only RCTs were included because they are a 'gold standard'. Although this may be true in certain circumstances, RCTs also often represent an unrealistic level of resource and investment in both intervention delivery and evaluation. There is a large body of research that questions whether RCTs are suitable for evaluations of complex public health interventions, and why effects found in RCTs do not consistently replicate in real-world/routine delivery. In addition, most physical activity programmes are not RCTs and are conducted in single-group, non-research designs. Although I understand the need for RCTs in building this corpus, particularly around parameters of effectiveness, the point that these might be delivered in unrealistic conditions should be acknowledged in the discussion. This point would fit well in the last paragraph before the strengths and limitations, where other points of this nature are highlighted.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: I have previously been seconded to the Centre for Behaviour Change at UCL, and currently work with colleagues there on a project for the Policy Research Unit.

Reviewer Expertise: I have extensive research expertise in assessing the content and effectiveness of physical activity interventions, and in behaviour change methodology more broadly.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 03 April 2025

<https://doi.org/10.21956/wellcomeopenres.26422.r121203>

© 2025 Downey J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



John Downey 

University of Plymouth, Plymouth, England, UK

Having looked at the revised manuscript and responses to my initial peer review comments, I am happy that the authors have now made explicit the information I needed to ease the concerns I had. For me, this was more to do with articulating key elements rather than actual rigour. I think the discussion is still quite light and am unable to comment on the AI processes. Nonetheless, I think it is a useful addition to the field, and I would welcome leaning on the teams experience for another project I am involved with exploring how AI was can learning from key documents. I would be happy to see this indexed.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 16 October 2024

<https://doi.org/10.21956/wellcomeopenres.23937.r102652>

© 2024 Downey J. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



John Downey 

University of Plymouth, Plymouth, England, UK

Thank you for your manuscript exploring the appropriateness of annotations from the HBCP to date when applied the physical activity literature. The on-going work is novel, has utility, and if it achieves its aim would continue to revolutionise behavioural science. To be accessible and acknowledge the current context, I suggest the below amendments to the current version.

The title lacks accessibility for the uninitiated decreasing its impact and potential relevance outside a certain audience. I would encourage the authors to alter the wording to increase engagement and decrease the esoteric nature of the work.

Within the abstract some background on why this is important would be useful, instead of quoting the HBCP. Given health behaviours are such an important area and the large potential of the ontology and AI to short cut and improve our understanding of the literature to date, I think some reference to these contributions is needed. A minor point also relates to the term 'evidence synthesis'. I think you could unpack that to provide the setting that I lay out in the initial text.

Within the introduction I would be conscious of using the realist mantra as I am unsure data from intervention studies provide enough mechanistic and contextual data to answer that question. Some transparency on what the AI could do, has done, and is hindered by would also be useful to set the scene. Likewise in the introduction I think the current issues with the behavioural science literature needs acknowledging and some scene setting on why annotations are needed and how it links to AI being able to answer key questions about interventions.

The methods are comprehensive, and I like the inclusion of supplementary material and reference to previously peer reviewed elements. It does however look pragmatic and is, as such, lacking detail/transparency on how some procedures were undertaken and key decisions made i.e the survey in phase 1 is not discussed and I am left thinking about construct validity, roll out, completion, and the actual thematic analysis process. In phase 2, if the manual didn't have an answer, where are these changes recorded?

In my opinion the discussion does not achieve its aim. A greater immersion in the established literature and consultation with key areas that corroborate, help explain key findings, or detail on how this work fills a gap is needed.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

No source data required

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: I am interested in the application of behavioural science by health professionals to support physical activity in those with long term conditions. More recently I am doing applied research exploring digital transformation in healthcare. I would consider myself a behavioural and implementation scientist with an interest in evaluation methodology in complex systems.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Mar 2025

Oscar Castro

Reviewer's comment: Thank you for your manuscript exploring the appropriateness of annotations from the HBCP to date when applied the physical activity literature. The on-going work is novel, has utility, and if it achieves its aim would continue to revolutionise behavioural science. To be accessible and acknowledge the current context, I suggest the below amendments to the current version.

Response: We are grateful for the time and energy you expended on our behalf. In the following sections, we will respond to each of your points and suggestions.

Reviewer's comment: *The title lacks accessibility for the uninitiated decreasing its impact and potential relevance outside a certain audience. I would encourage the authors to alter the wording to increase engagement and decrease the esoteric nature of the work.*

Response: We have rephrased the title to better catch the reader's attention and highlight the main study aim (i.e., evaluate the transferability of HBCP's ontology coding methods developed for smoking cessation interventions to the physical activity domain):

- Old title: Creating a body of physical activity evidence to test the generalisation of annotation methods for automated evidence synthesis.
- New title: From smoking cessation to physical activity: Can ontology-based methods for automated evidence synthesis generalise across behaviour change domains?

Reviewer's comment: *Within the abstract some background on why this is important would be useful, instead of quoting the HBCP. Given health behaviours are such an important area and the large potential of the ontology and AI to short cut and improve our understanding of the literature to date, I think some reference to these contributions is needed. A minor point also relates to the term 'evidence synthesis'. I think you could unpack that to provide the setting that I lay out in the initial text.*

Response: We have amended the background section of the abstract to strengthen the study rationale: "Developing behaviour change interventions able to tackle major challenges such as non-communicable diseases or climate change requires effective and efficient use of scientific evidence."

However, we are constrained in the abstract due to word count limits, so we have further addressed the reviewer's comments (i.e., unpack the term "evidence synthesis") in the introduction: "Evidence synthesis refers to the process of compiling relevant studies on a specific topic to answer..." (paragraph 2).

Reviewer's comment: *Within the introduction I would be conscious of using the realist mantra as I am unsure data from intervention studies provide enough mechanistic and contextual data to answer that question. Some transparency on what the AI could do, has done, and is hindered by would also be useful to set the scene. Likewise in the introduction I think the current issues with the behavioural science literature needs acknowledging and some scene setting on why annotations are needed and how it links to AI being able to answer key questions about interventions.*

Response: We have amended the introduction to explain the need for an approach such as the HBCP's and to expand on the role of annotations in the proposed evidence synthesis approach. See new paragraph 2:

"Evidence synthesis refers to the process of compiling relevant studies on a specific topic to answer a particular research question (Langlois et al., 2018). In recent decades, researchers have placed a stronger emphasis on rigor and accountability of evidence synthesis methods – given birth to the now-standard systematic review methodology – and developed new approaches (e.g., meta-analysis) to statistically combine results from multiple studies (Thomas, 2024). These advancements have undoubtedly benefited behavioural science and

other fields, for example, to help ground decisions on scientific evidence (Signore & Campagna, 2023). However, many argue that current evidence synthesis methods have inherent limitations (Moore et al., 2022; Siontis & Ioannidis, 2018) and fail to leverage recent advancements in information and computing sciences which could help address research questions more effectively and efficiently (Sharp et al., 2023; Michie & Johnston, 2017). Behaviour change evidence is currently extracted from study reports through a manual, lengthy, and error-prone process, without a shared conceptual and linguistic framework in the field to facilitate true cumulative knowledge. This leads to research waste (Roberts & Ker, 2015), as findings cannot be easily integrated with other research, and represents a missed opportunity to advance our understanding of behaviour change for the better."

In response to reviewer 1, we have also elaborated on the HBCP's information extraction approach and how it is positioned within the scope of existing work (introduction) and on the limitations of an AI-based evidence synthesis approach (discussion).

Reviewer's comment: *The methods are comprehensive, and I like the inclusion of supplementary material and reference to previously peer reviewed elements. It does however look pragmatic and is, as such, lacking detail/transparency on how some procedures were undertaken and key decisions made i.e the survey in phase 1 is not discussed and I am left thinking about construct validity, roll out, completion, and the actual thematic analysis process. In phase 2, if the manual didn't have an answer, where are these changes recorded?*

Response: We have added further details in phase 1 regarding the survey for physical activity experts (see paragraph below). To summarise:

- We have removed the word "thematic analysis" as it can be misleading to the reader. Responses from the experts were collated (merging some of them when similar feedback was provided) and then brought to the wider team for discussion.
- We have clarified the deployment (roll out) happened in 2018.
- There was no construct validity testing as the purpose of the survey was to gather expert feedback on our preliminary list of physical activity behavioural outcome codes (i.e., the survey did not aim to measure a specific concept or theoretical construct).

"Twenty-three international experts in physical activity research were invited to give feedback on the physical activity behavioural outcome codes resulting from the previous step. Experts included 14 behavioural scientists and public health stakeholders that had previously been invited to advise on the HBCP, and nine additional stakeholders identified by the project team. An online questionnaire was emailed to experts in 2018 using Qualtrics XM TM software (free alternatives include Google Forms or LimeSurvey; we chose Qualtrics as we considered it to be the superior option from a technical standpoint). The survey was designed to be completed within 20 minutes and was divided into six categories (full survey available as online supplementary material 1; West et al., 2023b): (...) Experts were asked whether they thought any codes should be changed or added within each category and, if so, which ones should be changed or added. The responses were collated by two researchers (EN & EH), with feedback combined where applicable (i.e., merging similar responses by different experts) and discussed internally by the research team. Revisions were made to the physical activity behavioural outcomes specified in the annotation code set."

In relation to the physical activity annotation manual, we have clarified in the main text that the changes performed to the document during the annotations: “were additive (i.e., we did not change the *modus operandi* but rather expanded the manual’s instructions with some edge cases identified during the annotations).”

We have also highlighted in the annotation manual (File 2 supplementary materials) the portions that were added as a result of this process. For example, during the annotations debriefs there were some uncertainties on how to annotate the number of participants included in studies that use “intention-to-treat”, so further clarifications were added to the manual in this regard (page 32).

Last, it is worth clarifying that the annotation stage resulted in changes to the manual but (most importantly) also resulted in changes to the physical activity annotation code set (e.g., Office facility was added to capture work-based physical activity interventions). These are reflected in the paper (results section, stage 4).

Reviewer’s comment: *In my opinion the discussion does not achieve its aim. A greater immersion in the established literature and consultation with key areas that corroborate, help explain key findings, or detail on how this work fills a gap is needed.*

Response: We have amended the first part of the discussion (paragraphs 1-3) to better emphasise the knowledge gap our study aims to fill and to include a wider range of literature. In total, six citations outside the “HBCP ecosystem” have been incorporated into the discussion. However, as indicated to reviewer 1 in relation to the introduction, we are not aware of any other efforts in behavioural science that seek to use ontologies for the automated extraction and synthesis of behaviour change interventions (we have confirmed this by conducting a rapid search through both Elicit and SCITE, using the prompt: “summarise initiatives that leverage ontologies for evidence synthesis in behavioural science”) and thus it is challenging to find analogous work to contrast our approach and findings.

Competing Interests: No competing interests were disclosed.

Reviewer Report 09 October 2024

<https://doi.org/10.21956/wellcomeopenres.23937.r100769>

© 2024 Gültzow T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Thomas Gültzow 

Open University of the Netherlands, Heerlen, The Netherlands

The current manuscript provides a valuable and detailed outline, along with a tutorial, on how to design a Human Behaviour-Change Project (HBCP) corpus. The manuscript is well-written and offers a robust framework for other scholars to follow in conducting similar work. However, there are several areas that could be refined to improve the clarity, transparency, and overall impact of the article.

Below are my key comments and suggestions for your consideration:

Major Comments

Citations and Scope: The citations are heavily focused on the same research team, which is understandable given that the work builds on previous outputs. However, this might create the impression of an isolated 'echo chamber' in which the work was conducted.

Suggestion: You may want to consider integrating more citations from external sources and related works outside of the core team to situate your research within the broader scientific landscape. Including more critical perspectives might also enrich the discussion.

Transparency in Methodology: While the openness in sharing materials is commendable, some elements of the research process remain somewhat vague, such as the exact search strings used. Also the work does not seem to be preregistered.

Suggestion: It may enhance the transparency of the manuscript to provide more detail on each aspect of the process. If proprietary software was used, including a brief rationale for choosing it over open alternatives could help clarify its use, even though the highlighting of open software for others is a great approach. It would also be helpful to provide clarity on whether preregistration was considered (or done), and if not, to explain the rationale behind this decision.

Inclusion of Sedentary Behaviour: The inclusion of sedentary behaviour in the study is not clearly justified from the outset. I have always understood that physical activity experts view sedentary behaviour as a separate, albeit related, behaviour to physical activity. Without a clear rationale, readers may be left unclear on its relevance to the study.

Suggestion: Offering a clear rationale for including sedentary behaviour early in the manuscript could improve reader understanding.

Geographical Representation of Experts: The experts involved seem to be *predominantly* based in English-speaking countries and the Global North.

Suggestion: If this was a conscious choice, or if it reflects a limitation, you might consider discussing how this geographical focus could affect the generalisability of your findings. This could also highlight potential avenues for broader global engagement in future work.

Diversity of the Knowledge Base: While the manuscript addresses the clarity of the intervention reports as a limitation, it overlooks other critical issues, such as the lack of diversity in the sources. This omission could pose significant challenges in the future, particularly when it comes to the effectiveness of interventions for marginalised populations.

Suggestion: Expanding on the limitations related to diversity, as well as other relevant factors beyond clarity, would help provide a more comprehensive discussion of the strengths and limitations of your work. This could also offer a clearer understanding of how these limitations might affect the generalisability of your findings.

Minor Comments

Erroneous Bracket: In the sentence: "The BCIO offers a set of definitions for describing key entities of behavioural interventions and the casual and semantic relationship between those entities, covering intervention content (Corker et al., 2022; Marques et al., 2023), engagement, population, setting (Norris et al., 2020), target behaviour (Schenk et al., 2024), mechanisms of action (Schenk et al., 2023), and delivery (including mode (Marques et al., 2020), source (Norris et al., 2021), style (Wright et al., 2023) and schedule)." there appears to be a misplaced bracket towards the end.

Suggestion: I would remove the bracket at the end of the sentence.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Not applicable

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Behavioral and decision-making science focusing on inclusivity, inequalities, and facilitating societal transitions.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 10 Mar 2025

Oscar Castro

Reviewer's comment: The current manuscript provides a valuable and detailed outline, along with a tutorial, on how to design a Human Behaviour-Change Project (HBCP) corpus. The manuscript is well-written and offers a robust framework for other scholars to follow in conducting similar work. However, there are several areas that could be refined to improve the clarity, transparency, and overall impact of the article.

Response: Thank you for your effort in reviewing the manuscript and for your constructive feedback. We answer to your queries below.

Reviewer's comment: *Below are my key comments and suggestions for your consideration:*
Major Comments

Citations and Scope: The citations are heavily focused on the same research team, which is understandable given that the work builds on previous outputs. However, this might create the impression of an isolated 'echo chamber' in which the work was conducted.

Suggestion: You may want to consider integrating more citations from external sources and related works outside of the core team to situate your research within the broader scientific landscape. Including more critical perspectives might also enrich the discussion.

Response: Aside from the Human Behaviour Change Project (HBCP), we are not aware of any other projects in behavioural science that aim to leverage domain-specific ontologies for the automated extraction and synthesis of behaviour change interventions (we have confirmed this by conducting a rapid search through both Elicit and SCITE, using the prompt: "summarise initiatives that leverage ontologies for evidence synthesis in behavioural science"). As a result, most citations refer to prior HBCP work.

However, we acknowledge there is a wider body of literature on automated information extraction of behavioural science research that does not use an ontology-based approach. We have briefly addressed this in the introduction to better contextualise the HBCP within the broader research landscape:

"(...). This ontology-based approach distinguishes the HBCP from other information extraction efforts in behavioural and social sciences, which typically rely on different approaches such as distant supervision (e.g., Wei et al., 2022), classifiers (e.g., Zielinski & Mutschke, 2017), or neural networks (e.g., Shen et al., 2022) and are mostly focused on extracting data from abstracts (Legate et al., 2024; Schmidt et al., 2023), likely due to their greater availability and ease of access compared to full texts."

In response to reviewer 2 we have also expanded the introduction to provide a stronger rationale for the HBCP approach and this led to the inclusion of several citations beyond the HBCP (e.g., see new paragraph 2). In total, 12 citations outside the "HBCP ecosystem" have been incorporated into the introduction.

Reviewer's comment: *Transparency in Methodology: While the openness in sharing materials is commendable, some elements of the research process remain somewhat vague, such as the exact search strings used. Also the work does not seem to be preregistered.*

Suggestion: It may enhance the transparency of the manuscript to provide more detail on each aspect of the process. If proprietary software was used, including a brief rationale for choosing it over open alternatives could help clarify its use, even though the highlighting of open software for others is a great approach. It would also be helpful to provide clarity on whether preregistration was considered (or done), and if not, to explain the rationale behind this decision.

Response: The keywords stated in the manuscript are all the terms used during the search.

We have clarified this in the revised version and also specified the level in which these terms were searched (i.e., title and abstract):

"(...). The search was performed on 20.01.2021 and used the following search string at the title or abstract level: MVPA or "moderate-to-vigorous physical activity" or MPA or VPA or "moderate physical activity" or "vigorous physical activity" or "strenuous physical activity" or "hard physical activity", with reports additionally filtered using the Microsoft Academic's built-in Randomised Controlled Trial classifier."

It is worth noting that our search was different from a traditional systematic review search. Microsoft Graph does not work like a standard academic database such as PubMed or Web of Science (e.g., note the use of the RCT in-built classifier (filter) in combination with the search terms) and our goal was to generate an extensive (but not comprehensive) subsample of RCTs on physical activity behaviour change to serve as a training set for the Knowledge System.

We have added further details in the manuscript to justify the use of the proprietary software, including:

- EPPI-Reviewer: "(...). We used EPPI-Reviewer because the research team had already expertise and training with this software and the developers were willing to change it to support our needs at no additional cost.
- Qualtrics: "(...). An online questionnaire was emailed to experts in 2018 using Qualtrics XM TM software (free alternatives include Google Forms or LimeSurvey; we chose Qualtrics as we considered it to be the superior option from a technical standpoint)."

Regarding pre-registration, given the descriptive nature of this work we deemed it as not necessary (i.e., there is no hypothesis or analyses reported which could be susceptible to p-hacking or HARKing). However, we recognise that pre-registration could still have been valuable (e.g., to aid discoverability and avoid duplication of research efforts) and we have included this as a study limitation:

"(...). Last, while this study was not pre-registered due to its largely descriptive nature – lacking hypotheses or analyses susceptible to p-hacking or other questionable research practices – we acknowledge that pre-registration could still have been beneficial (e.g., to enhance discoverability and help prevent unintentional duplication of research efforts). Future hypothesis-driven research conducted using this published corpus, such as training and evaluating information extraction algorithms, should be pre-registered to ensure transparency and minimise publication bias."

Reviewer's comment: *Inclusion of Sedentary Behaviour: The inclusion of sedentary behaviour in the study is not clearly justified from the outset. I have always understood that physical activity experts view sedentary behaviour as a separate, albeit related, behaviour to physical activity. Without a clear rationale, readers may be left unclear on its relevance to the study. Suggestion: Offering a clear rationale for including sedentary behaviour early in the manuscript could improve reader understanding.*

Response: We have provided a stronger rationale for the inclusion of sedentary behaviour, based on the fact that both outcomes are typically measured and reported in physical activity studies:

"(...). We included reviews on sedentary behaviour as part of the scoping work because, while sedentary behaviour is a distinct behaviour within the domain of physical activity, they form part of the same energy expenditure continuum and in many studies – particularly those using accelerometers – both are reported."

Reviewer's comment: *Geographical Representation of Experts: The experts involved seem to be predominantly based in English-speaking countries and the Global North.*

Suggestion: If this was a conscious choice, or if it reflects a limitation, you might consider discussing how this geographical focus could affect the generalisability of your findings. This could also highlight potential avenues for broader global engagement in future work.

Response: We have added this as a limitation in the discussion section and elaborated on how we plan to mitigate geographical bias in upcoming work:

"Another limitation of this study is the geographical concentration of physical activity experts consulted, primarily from English-speaking countries and the Global North. This focus may limit the generalisability of the findings, as perspectives and contexts from other regions may not be fully represented. As part of the upcoming Advancing Prevention Research In Cancer through Ontology Tools (APRICOT) project – which extends the HBCP work by developing tools to facilitate the use of ontologies in behavioural science (Michie et al., 2024) – we plan to set up a diverse Community of Practice to ensure broader global engagement in future work."

Reviewer's comment: *Diversity of the Knowledge Base: While the manuscript addresses the clarity of the intervention reports as a limitation, it overlooks other critical issues, such as the lack of diversity in the sources. This omission could pose significant challenges in the future, particularly when it comes to the effectiveness of interventions for marginalised populations.*

Suggestion: Expanding on the limitations related to diversity, as well as other relevant factors beyond clarity, would help provide a more comprehensive discussion of the strengths and limitations of your work. This could also offer a clearer understanding of how these limitations might affect the generalisability of your findings.

Response: We agree with this point and that it has not been sufficiently elaborated in the initial version. Despite its potential advantages, an ontology-based evidence synthesis approach will only be as good – or as bad – as the underlying evidence, which is not free of bias. This is a limitation shared with systematic reviews and other "traditional" evidence synthesis methods which potential users should be aware of. We have added the following text and citations to the discussion section:

"Last, it is also important to recognise that artificial intelligence systems are only as good as the data they are trained on and operate with. The available evidence on behaviour change is not free of bias. For example, successful interventions are more likely to be reported and

published (Nissen et al., 2016). Similarly, most behavioural research is conducted in high-income countries (Miranda et al., 2010), with predominantly white samples (Oh et al., 2015). Thus, findings might not be applicable to other contexts and ethnic groups."

Reviewer's comment: Minor Comments

Erroneous Bracket: In the sentence: "The BCIO offers a set of definitions for describing key entities of behavioural interventions and the casual and semantic relationship between those entities, covering intervention content (Corker et al., 2022; Marques et al., 2023), engagement, population, setting (Norris et al., 2020), target behaviour (Schenk et al., 2024), mechanisms of action (Schenk et al., 2023), and delivery (including mode (Marques et al., 2020), source (Norris et al., 2021), style (Wright et al., 2023) and schedule)." there appears to be a misplaced bracket towards the end. Suggestion: I would remove the bracket at the end of the sentence.

Response: We have removed the bracket and slightly rephrased the sentence for clarity.

Competing Interests: No competing interests were disclosed.