The use of directed acyclic graphs (DAGs) in physical activity and nutrition research: A scoping review of the literature

Gabriel P Esteves[1,2], Lucas Porto Santos[1,2], Joanna Shim[3], Jake Slaton[3], Eimear Dolan[1,2,4], Paul Swinton[3]

[1] Applied Physiology and Nutrition Research Group – School of Physical Education and Sport and Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Sao Paulo, SP, Brazil

[2] Center of Lifestyle Medicine; Laboratory of Assessment and Conditioning in Rheumatology; Hospital das Clínicas HCFMUSP, Faculdade de Medicina FMUSP, Universidade de Sao Paulo, Av. Dr. Arnaldo, 455, 3° andar, São Paulo, SP 01246-903, Brazil

[3] School of Health, Robert Gordon University, Garthdee Road, Aberdeen, UK

[4] School of Arts, Sciences and Humanities, University of Sao Paulo, Sao Paulo, Brazil.

Correspondence to: Paul A. Swinton, Robert Gordon University, School of Health, Aberdeen, UK. Email: p.swinton@rgu.ac.uk

Preprint submitted to SportRxiv – not yet peer reviewed.

Please cite as: Esteves et al. (2025) The use of directed acyclic graphs (DAGs) in physical activity and nutrition research: A scoping review of the literature.

Abstract

Introduction: Directed acyclic graphs (DAGs) illustrate causal structures, but their application in physical activity and nutrition research remains unclear. We aimed to characterise DAG use in this literature, highlighting best practices and areas for improvement.

Methods: We conducted a scoping review of DAG use in physical activity and nutrition-related articles published between 1999 and 2024, extracting data on study topic, design, DAG justification and construction, number of arcs, nodes, exposures, outcomes, confounders, mediators, mediator-outcome confounders, competing exposures, and instrumental variables.

Results: Of 115 included studies, 110 contained extractable DAG data. Five could not be extracted due to DAG size or unfixed nodes. Among the 110 studies, 86 (78%) made their DAG available. Most (61, 55%) did not specify methods for identifying variables or causal arcs. When specified, the most common approach was literature review (32, 29%). DAGitty software was used in 68 studies (62%). A total of 96 DAGs were identified, with the majority addressing nutritional exposures (75, 68%). DAGs had a median number of 13 nodes; 2 causal paths; 6 confounders; 1 mediator; and 0 mediator-outcome confounders, instrumental variables and competing exposures.

Conclusion: DAGs support causal inference but their value depends on accurately representing the true causal structure. Many studies lacked a systematic approach for DAG construction and omitted potentially informative nodes such as mediators and mediator-outcome confounders. We provide recommendations to improve the use and transparency of DAGs in physical activity and nutrition research.

Keywords: causal inference; confounding; collider bias; epidemiology; kinesiology.

Introduction

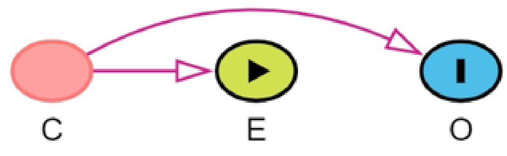
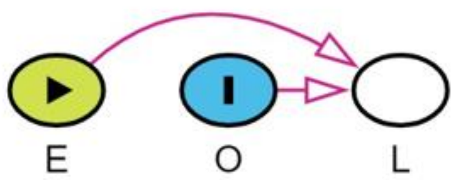
Establishing cause-effect relationships, that is, whether a specific treatment, intervention, or condition under investigation directly affects a given outcome (1), is fundamental to advancing physical activity and nutrition research. Examples of causal questions include “does exercise training reduce the risk of bone fractures?” or “does consuming more ultra-processed foods increase cardiovascular risk?”. In the health sciences, randomized controlled trials (RCTs) are commonly employed to determine cause-effect relationships (2); however, in many cases, randomized experiments are not feasible or ethical to conduct. For example, it may not be ethical to assign individuals to consume diets high in saturated fats or to restrict their physical activity levels for prolonged periods to determine the long-term influence of these practices on health outcomes. Furthermore, both nutrition and physical activity are complex behaviours that influence each other, along with multiple other outcomes, such as stress levels or sleep patterns. Understanding these complex relationships is fundamental to developing fit-for-purpose interventions, but controlling these parameters in free-living situations and for prolonged periods is difficult. As such, many relevant scientific questions in physical activity and nutrition research—particularly those requiring large samples and long follow-up periods—may be best addressed using observational designs, such as prospective cohort studies.

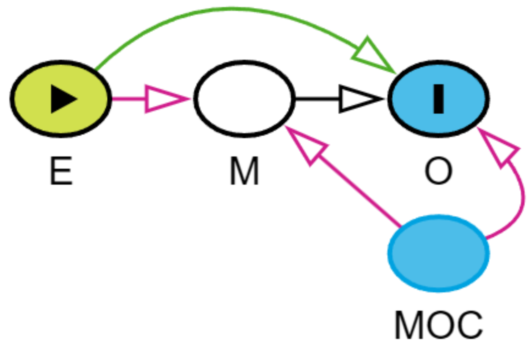
Although powerful tools with considerable potential to advance understanding of how nutritional and physical activity behaviours influence health outcomes, observational designs are vulnerable to various sources of bias that can limit their interpretation (3). These biases can be broadly classified within two categories, namely those arising from common causes (generally referred to as confounding) and those arising from conditioning on common effects (also known as collider bias) (4). Examples of how these biases may manifest in nutrition and physical activity-based investigations are described in Table 1.

To reduce these biases and strengthen causal inference, researchers are advised to first define and visualize assumed causal relationships between variables of interest (5). This can be achieved by constructing a directed acyclic graph (DAG), whereby the hypothesized direction of causal relationships between variables of interest are plotted (5,6). DAGs help identify potential sources of confounding and collider bias, guiding the selection of appropriate statistical models to reduce bias and improve causal inference (7).

A previous systematic review that assessed DAG use in applied health research showed that their use is increasing in this area (6). Whether this increase is reflected in the more specific areas of physical activity and nutrition remains unclear. It is also unclear whether DAGs are being used effectively, as these fields share unique methodological challenges that can foster biases. For example, researchers in physical activity research may inadvertently adjust for variables that are not confounders, but rather lay on the causal path between exposure and outcome (*i.e.,* mediators), therefore underestimating the total effect. A further common challenge across both nutrition and physical activity research is measurement error. Indirect tools such as dietary and physical activity surveys are prone to misreporting and information bias—issues that DAGs can help visualize and potentially mitigate (8). Therefore, the aim of this scoping review is to map how DAGs are currently being used in nutrition and physical activity sciences, including the justifications provided for their use, the adequacy of their design, and their practical application in these fields.

Table 1. Definitions and practical examples of bias in physical activity and nutrition research

Bias type (also known as)	**Definition**	**Example of bias in physical activity and nutrition research**	**DAG depiction**
Bias due to common causes (confounding)	When the exposure and outcome share a common cause, creating a non-causal association between them.	Example 1) A cross-sectional study examining the effect of vitamin D levels (E) on osteoporosis incidence (O) fails to account for physical activity (C), which influences both vitamin D and osteoporosis.	
Bias due to conditioning on common effects (selection bias, collider bias).	When the exposure and outcome both influence a third variable (a common effect), conditioning on this variable (*e.g.*, through selection or adjustment) induces a spurious association between exposure and outcome (Example 2). Collider bias can also occur	Example 2) A prospective cohort study examining the effect of exercise habits (E) on muscle functionality (O) experiences greater loss to follow-up among non-exercising individuals. Because muscle function affects likelihood of follow-up (those with better muscle function are more likely to remain), analysing only those that remain (L: follow-up availability) conditions on a	

	when adjusting for a mediator that shares a common cause with the outcome (mediator-outcome confounding, Example 3).	collider. This biases the association between exercise and muscle function because the non-exercisers who remain in the study tend to have better muscle function than non-exercisers in the full population.	
		Example 3) A prospective cohort study estimating the direct effect of physical activity levels (E) on total mortality adjusts for blood pressure (M), a mediator. However, smoking (MOC) influences both blood pressure and total mortality, introducing bias through mediator-outcome confounding.	

E: exposure; O: outcome; C: common cause of exposure and outcome (confounder); L: common effect of exposure and outcome (collider); M: mediator; MOC: mediator-outcome confounder. White nodes denote variables that have been conditioned on; purple arrows indicate open biasing paths. Figures created using DAGitty.com.

Methods

The review was conducted in accordance with the JBI methodology for scoping reviews (9).

Eligibility criteria

Inclusion criteria for this review were defined using the PCC (Population or Participants, Concept, Context) framework (9). The concept of interest was the use of DAGs to inform research investigating causal phenomena, and the context physical activity and nutrition-related research. For the purposes of this review, physical activity research was defined as studies with any physical activity-related exposure of interest, such as physical activity levels, participation in exercise or rehabilitation programs, measurements of muscle strength or function, and related factors; while nutrition research was defined as studies with a nutrition-related exposure of interest, such as dietary patterns, specific nutrients (*e.g.*, iron intake), and similar factors. Studies were restricted to human populations.

Types of Sources

Peer-reviewed studies of any research design that reported employing a DAG as part of its investigation (*i.e.*, prospective and retrospective cohort studies, case-control studies, cross-sectional studies, RCTs, meta-analyses) were included. Reviews and tutorial papers were excluded.

Search strategy

An initial limited search of MEDLINE was performed to identify relevant articles and refine the main search strategy. Subsequently, MEDLINE, Embase, Cochrane Central and SPORTDiscus were searched from January 1999—following the publication of a seminal paper on DAGs (5)—to the present (search conducted in March 2024), using the terms ‘graphical model theory’, ‘directed acyclic graph(s)’, ‘causal diagram(s)’, ‘causal graph(s)’ or ‘causal DAG’ (6) concatenated with identifiers for the fields of physical activity and nutrition research (*e.g.*, ‘physical activity’, ‘exercise’, ‘nutrition’, ‘diet’). The initial search string was developed for MEDLINE, and was then translated for the remaining databases utilizing Polyglot Search Translator (10). The detailed search strings used are available in Supplementary File 1. Searches were conducted across all available fields (title, abstract, keywords, and index terms).

Study/Source of Evidence selection

All retrieved citations were uploaded into Covidence (Veritas Health Innovation, Melbourne, Australia), an online systematic review screening tool, which was used to remove duplicates and manage screening. Each title and abstract were screened independently by G. Esteves and J. Shim against the inclusion criteria. Full texts were retrieved and again screened independently by G. Esteves and J. Shim. Reasons for exclusion at full text were recorded. Discrepancies between reviewers were resolved through discussion or by consulting a third reviewer (P. Swinton). The search results are reported using the PRISMA-ScR flow diagram (11).

Definitions used for review

To more clearly describe the data extracted from the identified papers, we first offer some operational definitions of important terms related to DAGs according to their common use in the literature (6,12–14). DAGs are diagrams that represent the data-generating process. They are *directed* in the sense that each variable, represented by a node, is connected to other variables by arrows (arcs) that assume a single direction. A *path* is the sequence of arcs that connect one variable to another. DAGs are also *acyclic*, meaning that a variable cannot cause itself.

Exposure variables are the main cause under study, that affects a given *outcome*. For example, a DAG on how dietary protein intake influences bone mineral density would consider protein as the exposure, whereas bone mineral density would be the outcome. Each DAG typically focuses on one main relationship of interest (*i.e.*, includes one exposure and outcome). This relationship is affected by variables which can bias interpretation of the true relationship between the exposure and outcome. In this review, we focused on confounders, mediators, mediator-outcome confounders, instrumental variables, and competing exposures.

Confounders are those variables that have arcs pointed at both the exposure and the outcome, *i.e.*, they have a causal effect on both variables. For instance, age may a confounder when considering the effect of protein on bone, considering that different age groups might have differing protein intakes, and that age also affects bone mineral density directly. Once potential confounders have been identified, they can then be *conditioned on* using statistical approaches, *e.g.*, via covariate adjustment in multivariable regression. This adjustment closes the biased path between exposure and outcome introduced by the confounder (Figure 1, panel A). *Mediators* are variables that lie on the causal path between the exposure and the outcome, with part of the causal effect of the exposure acting through the mediator. For example, in a DAG where physical activity influences cardiovascular mortality both directly and indirectly via blood pressure, blood pressure is a mediator, meaning it explains part (but not all) of the effect of physical activity on cardiovascular mortality. Studies may wish to estimate the total effect of the intervention, where the estimated effect of the exposure includes that of the mediator; or, they may wish to estimate the direct effect of the exposure, where the effect of the mediator is separated from that of the exposure.

Mediator-outcome confounders are variables that causally influence both a mediator and the outcome. They do not bias estimates of the **total effect**, as this does not involve conditioning on the mediator. However, when researchers aim to partition the total effect into **direct** and **indirect** components, they must condition on the mediator. In doing so, the mediator acts as a collider between the exposure and the mediator–outcome confounder. If the mediator–outcome confounder is not also adjusted for, this opens a non-causal path from exposure to outcome, introducing bias. To estimate the direct effect without bias, both the mediator and the mediator–outcome confounder must be included in the adjustment set (Figure 1, panel B).

Instrumental variables are those that only affect the exposure and are independent from confounders, while *competing exposures* are those that only affect the outcome, and are independent from the exposure (Figure 1, panel C). Instrumental variables can be used in mendelian randomization studies in an attempt to identify causal effects in observational data. For instance, studies have used single-nucleotide polymorphisms associated with physical activity to estimate its causal effect on depression (15). Competing exposure variables can potentially increase precision in estimating the outcome. For example, a family history of heart disease might affect cardiovascular mortality, while being independent from the exposure physical activity, potentially providing increased accuracy if conditioned on.

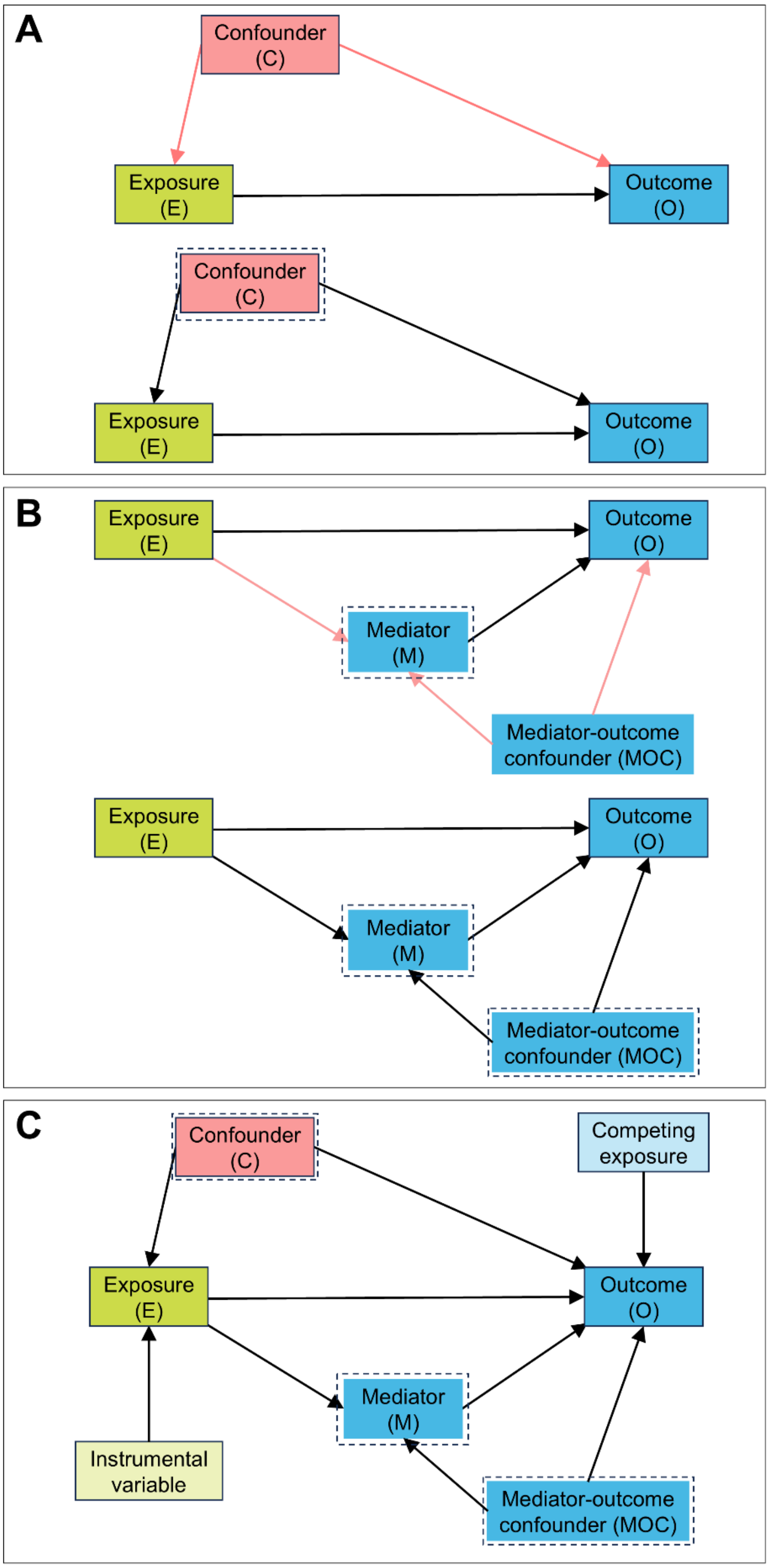


Figure 1. Illustration of main DAG components, such as nodes and paths. A) displays a confounder variable creating a biasing path (pink) between exposure and outcome. Conditioning (square, dashed box) on the variable closes the biasing path. B) displays an example of collider bias when adjusting for a mediator in the presence of a mediator-outcome confounder. C) displays all variables of interest extracted during the review.

After a DAG has been drawn, an *adjustment set* of variables can be identified according to the principles of graph model theory underlying DAGs (16). An adjustment set is the set of variables that, given the assumptions encoded in the DAG, will yield an unbiased estimate of the causal effect of the exposure on the outcome when adequately conditioned on. Knowing which variables do not need to be conditioned on is also relevant, as it helps researchers to avoid adjusting for unnecessary variables, which could inadvertently introduce noise or additional bias. This adjustment set can be obtained visually by applying the rules briefly summarised here or though software-based algorithms that apply these rules automatically, such as DAGitty (12).

Data Extraction

Two reviewers (G. Esteves and J. Slaton) extracted data from the included papers. Extracted variables covered three domains: study characteristics, DAG-related information, and statistics-related information. Study characteristics included first author; journal; publication year; study title and design (case-control, causal model creation, longitudinal cohort, mendelian randomization, randomized controlled trial, cross-sectional, systematic review and/or meta-analysis); research area (biomechanics, clinical nutrition, exercise epidemiology, exercise physiology, nutritional epidemiology, sports medicine, strength and conditioning); the database utilized for the analysis (name of study, cohort or secondary database); and sample size.

DAG-related information included whether the DAG was available in the publication; number of DAGs per paper; software used to visually construct the DAG; the development method (categorized as not reported, literature-based, expert-based, literature and expert-based, causal discovery algorithm or Delphi consensus); the stated justification for DAG usage (verbatim statements from authors); number of variables in the DAG; shortest and longest path between exposure and outcome; whether the DAG estimated the total effect, direct effect or both; number of edges; and whether the DAG adhered to the acyclic property. Additionally, we noted the name and frequency (counts) of each node type, namely, the number of exposures, outcomes, confounders, mediators, mediator-outcome confounders, instrumental variables, and competing exposures. We also computed whether or not each DAG included at least one of each type (*e.g.,* includes at least one confounder). To determine the name and frequency of each variable type, we initially registered the name of all variables verbatim as used by the original authors in their DAG. Then, these extracted names were recoded into overarching constructs to allow for meaningful aggregation. For instance, "diet", "food consumption" and "nutritional patterns" were all recoded as the overarching term "diet". Similarly, "physical activity", "leisure-time physical activity" and "exercise", were recoded as "physical activity". This recoding facilitated consistent counting across studies. Supplementary Table S1 details how each variable was recoded, and further explanation about the extraction and coding process is provided in Supplementary File 2.

Statistical method information included the type of model used (*e.g.,* generalized linear model, proportional hazards model); the functional form of the exposure–outcome association (*e.g.,* categorical, continuous linear, spline); whether the adjustment set was reported in some manner; and whether causal or associational language was used to interpret results. To classify the type of language, we reviewed the results and conclusion sections of each article. Statements such as "The exposure was associated with a change in the outcome" were classified as associational, whereas "The exposure led to a change in the outcome" or "The exposure had an effect on the outcome" were classified as causal. If both types of statements were present, the most emphasized and frequent type was recorded. Extracted quantitative data were summarised using medians and interquartile ranges for continuous variables, and counts and proportions for categorical variables.

Results

Figure 2 shows a flow diagram of study screening. Initially, 648 studies were identified through the literature search, and 2 additional studies were manually added. Following screening, a total of 115 studies were included in the review, with DAG-related data extracted from 110. These data were not extracted from 5 studies, either because the DAGs presented were too large (17,18), or because they lacked fixed variables (*e.g.,* the DAG served as a base model for other researchers rather than representing a fixed causal structure) (19–21). However, their article-related characteristics and DAG construction methods were still recorded and analyzed.

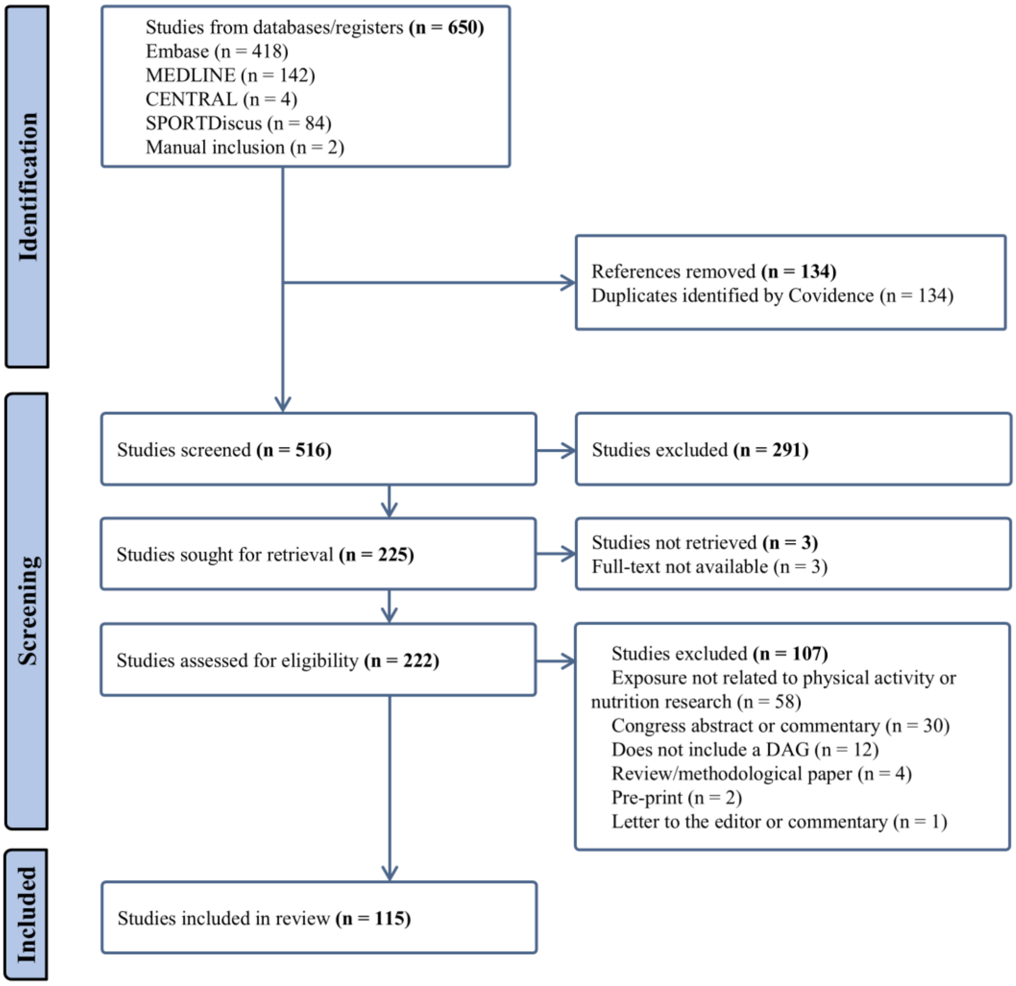


Figure 2. PRISMA-ScR flow diagram showing study identification and screening

Article characteristics

Figure 3 shows article-related characteristics of the included studies, namely the first author's country affiliation, research area, study design, and year of publication. A detailed description of all article characteristics is provided in Supplementary Table S2. Briefly, out of 115 articles, most first-authors were affiliated with institutions in Brazil (n=19, 17%) and China (n=18, 16%). The most common study areas were nutritional epidemiology (n=70, 61%) and exercise epidemiology (n=23, 20%), and most studies employed cross-sectional (n=52, 45%) or prospective cohort (n=50, 43%) designs. The earliest studies were published in 2013, and the number of published studies appeared to increase yearly (Figure 3, panel D), with most publications appearing in 2023 (n=30, 26%).

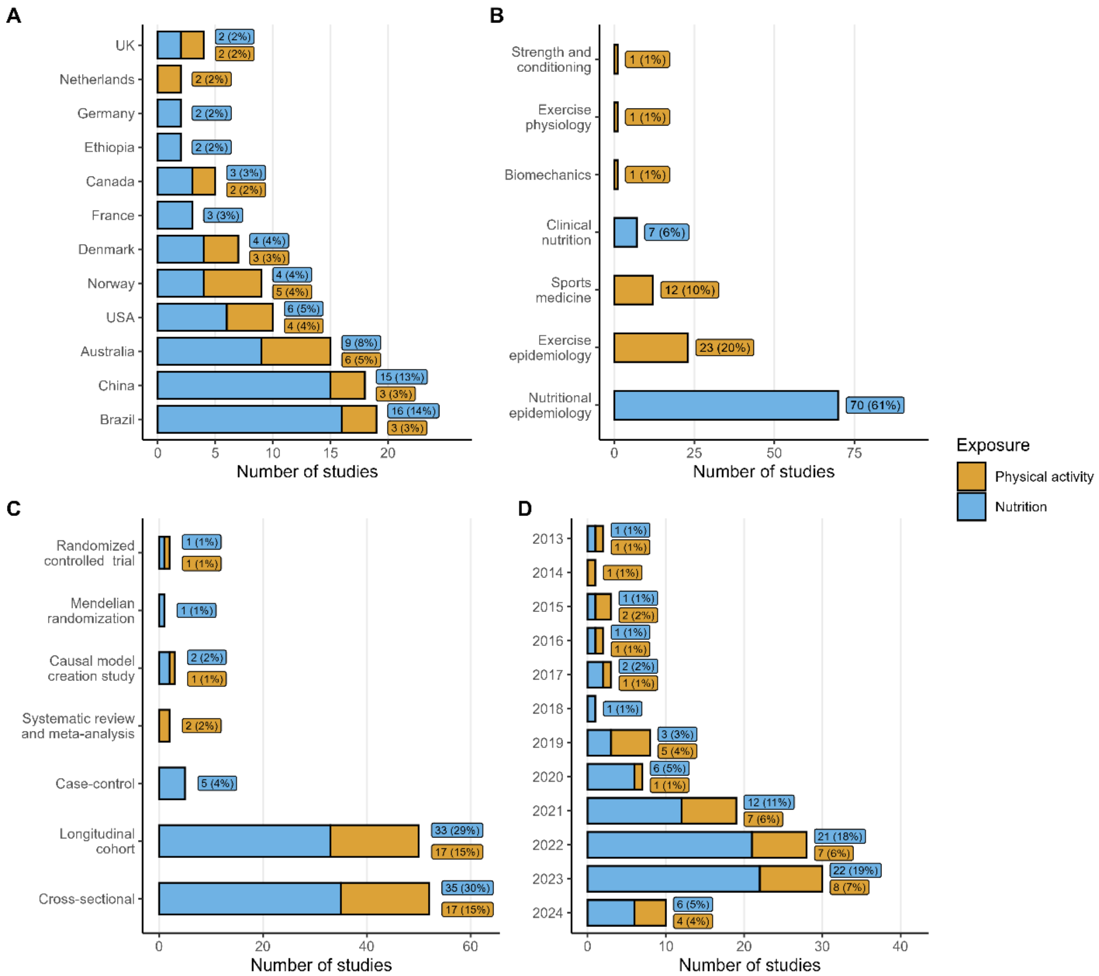


Figure 3. Article-related characteristics of included studies. Figure shows counts and percentages for articles according to A) country affiliation of the first-author; B) research area; C) study design; D) articles identified in each year. In panel A, countries with only one identified paper were omitted to facilitate visualization.

DAG-related characteristics

Table 2 presents key DAG-related information. Of the 110 studies for which DAG-related data were extracted, 86 (78%) made their DAG available. Approximately two-thirds of the studies examined nutrition-related exposures (n=75, 68%), while one-third examined physical activity-related exposures (n=35, 32%). The most frequently cited reason for constructing DAGs was covariate selection, with studies commonly stating they were used "*to identify covariates*", "*to select adjustment sets*" or "*to avoid incomplete adjustment or overadjustment*". Where software was specified, DAGitty (12) was the only tool reported (68, 62%). Most studies (n=61, 55%) did not describe a specific approach for DAG construction. Among those that did, the most common approach was literature-based (*i.e.*, formal or informal literature reviews) (n=32, 29%), and often described briefly, for example, that the DAG was "*based on the literature*", "*based on the recent literature and research evidence*" or "*based on a review of the literature*". However, some studies provided more detailed descriptions, such as those that conducted a systematic review to inform DAG construction (17,22,23), or supplied decision logs explaining each node and arc inclusion in the DAG and its supporting literature (20,24–26). Some studies (17,23,24) reported following the Evidence Synthesis for Constructing Directed Acyclic Graphs (ESC-DAGs) method (13). No studies reported using a Delphi consensus approach. The majority of studies (n=92, 84%) primarily used associative rather than causal language.

Among the 86 studies with available DAGs, most (n=77, 90%) presented a single DAG, with a total of 96 DAGs extracted. The median (IQR) number of nodes per DAG was 13 (9, 16), with 2 (1, 4) causal paths, 6 (3, 9) confounders, 1 (0, 3) mediator, 0 (0, 1) mediator-outcome confounders, and 0 (0, 0) instrumental variables or competing exposures. Most DAGs were acyclic (n=87, 91%) and comprised a median 32 (18, 49) edges.

Table 2. DAG-related characteristics in physical activity and nutrition research articles

Characteristic	Overall, N = 110	Physical activity, N = 35	Nutrition, N = 75
DAG availability	86 (78%)	24 (69%)	62 (83%)
DAG software			
DAGitty	68 (62%)	19 (54%)	49 (65%)
Not reported	42 (38%)	16 (46%)	26 (35%)
DAG development method			
Not reported	61 (55%)	19 (54%)	42 (56%)
Literature-based	32 (29%)	9 (26%)	23 (31%)
Expert-based	9 (8.2%)	2 (5.7%)	7 (9.3%)
Literature and expert-based	8 (7.3%)	5 (14%)	3 (4.0%)
Number of DAGs per paper			
One	77 (90%)	21 (88%)	56 (90%)
Two	8 (9.3%)	3 (12%)	5 (8.1%)
Three	1 (1.2%)	0 (0%)	1 (1.6%)
Language used			
Causal	18 (16%)	9 (26%)	9 (12%)
Associational	92 (84%)	26 (74%)	66 (88%)
Reported adjustment set	109 (99%)	35 (100%)	74 (99%)
Characteristic	**Overall, N = 96**	**Physical activity, N = 27**	**Nutrition, N = 69**
Nodes	13.0 (9.0, 16.0)	10.0 (6.0, 13.0)	14.0 (11.0, 18.0)
Causal paths	2.0 (1.0, 4.0)	2.0 (1.0, 3.0)	2.0 (1.0, 4.0)
Direct (Yes/No)	84 (88%)	23 (85%)	61 (88%)
Shortest path	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)	1.0 (1.0, 1.0)
Longest path	2.0 (1.0, 3.0)	2.0 (1.0, 2.0)	2.0 (1.0, 3.0)
Confounders	6.0 (3.0, 9.0)	4.0 (3.0, 8.0)	6.0 (4.0, 9.0)
Mediators	1.0 (0.0, 3.0)	1.0 (0.0, 2.5)	2.0 (0.0, 3.0)
Mediator-outcome confounder	0.0 (0.0, 1.0)	0.0 (0.0, 0.0)	0.0 (0.0, 1.0)
Competing exposures	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 1.0)
Instrumental variables	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)	0.0 (0.0, 0.0)
Included confounders	93 (97%)	26 (96%)	67 (97%)
Included mediators	58 (60%)	15 (56%)	43 (62%)
Included mediator-outcome confounders	27 (28%)	5 (19%)	22 (32%)
Included instrumental variables	22 (23%)	2 (7.4%)	20 (29%)
Included competing exposures	11 (11%)	2 (7.4%)	9 (13%)
Estimating total or direct effect			
Total effect	73 (76%)	21 (78%)	52 (75%)
Direct effect	14 (15%)	3 (11%)	11 (16%)
Both	9 (9.4%)	3 (11%)	6 (8.7%)
Number of edges	32 (18, 48)	22 (11, 33)	35 (20, 50)
DAG acyclic (Yes)	87 (91%)	25 (93%)	62 (90%)

Numerical data are reported as median (interquartile range lower and upper limits); categorical data are reported as n (%).

Commonly reported variables

Figure 4 provides a visual summary of the most commonly reported variables across node types. The most frequent exposures were diet (n=17, 18%), physical activity (n=13, 13%), breastfeeding, dietary inflammatory index, food security, structured exercise, sugar, and ultra-processed foods (all n=4, 4.1%, see Supplementary Table S3 for a comprehensive list). The most frequent outcomes were cancer (n=6, 6.1%), mental health (n=6, 6.1%), cardiometabolic health, diabetes or glucose metabolism status, and physical function/strength (all n=5, 5.1%), as well as obesity or weight status (n=4, 4%), among others. Additionally, 93 (97%) DAGs included confounders (Table 2), the most common being age (n=59, 9.8%), sex (n=48, 8.0%), education (n=34, 5.6%), smoking (n=34, 5.6%), income (n=26, 4.3%), and physical activity (n=17, 2.8%), followed by other less common factors related to demographic information or disease history (Supplementary Table S3). Mediators were present in 58 (60%) DAGs, the most frequent being BMI (n=23, 13%), cardiovascular disease, diabetes, and energy intake (all n=10, 5.8%), dyslipidaemia (n=9, 5.2%), adiposity or obesity (n=8, 4.6%), as well as physical activity (n=5, 2.9%). Mediator-outcome confounders were included in 27 (28%) DAGs, the most frequent being alcohol, physical activity and smoking (all n=8, 12%). Finally, 22 (23%) DAGs included instrumental variables, and 11 (11%) included competing exposures.

Statistical approaches

Out of the 115 studies, most relied on generalized linear models using linear (n=32, 28%) or logistic (n=32, 28%) approaches, followed by proportional hazards regression (cox regression, n=16, 14%), poisson family (n=8, 7%), or other variations. In 75 (67%) studies, the exposure-outcome relationship was treated as categorical, whereas 25 (22%) used linear continuous coding, and 8 (7.1%) used restricted cubic splines. Most analyses (n=73, 76%) estimated the total effect of the exposure rather than the direct effect (n=14, 15%) or both (n=9, 9.4%). Almost all studies reported their adjustment set in some format (n=109, 99%). Further methodological details are provided in Supplementary Table S4.

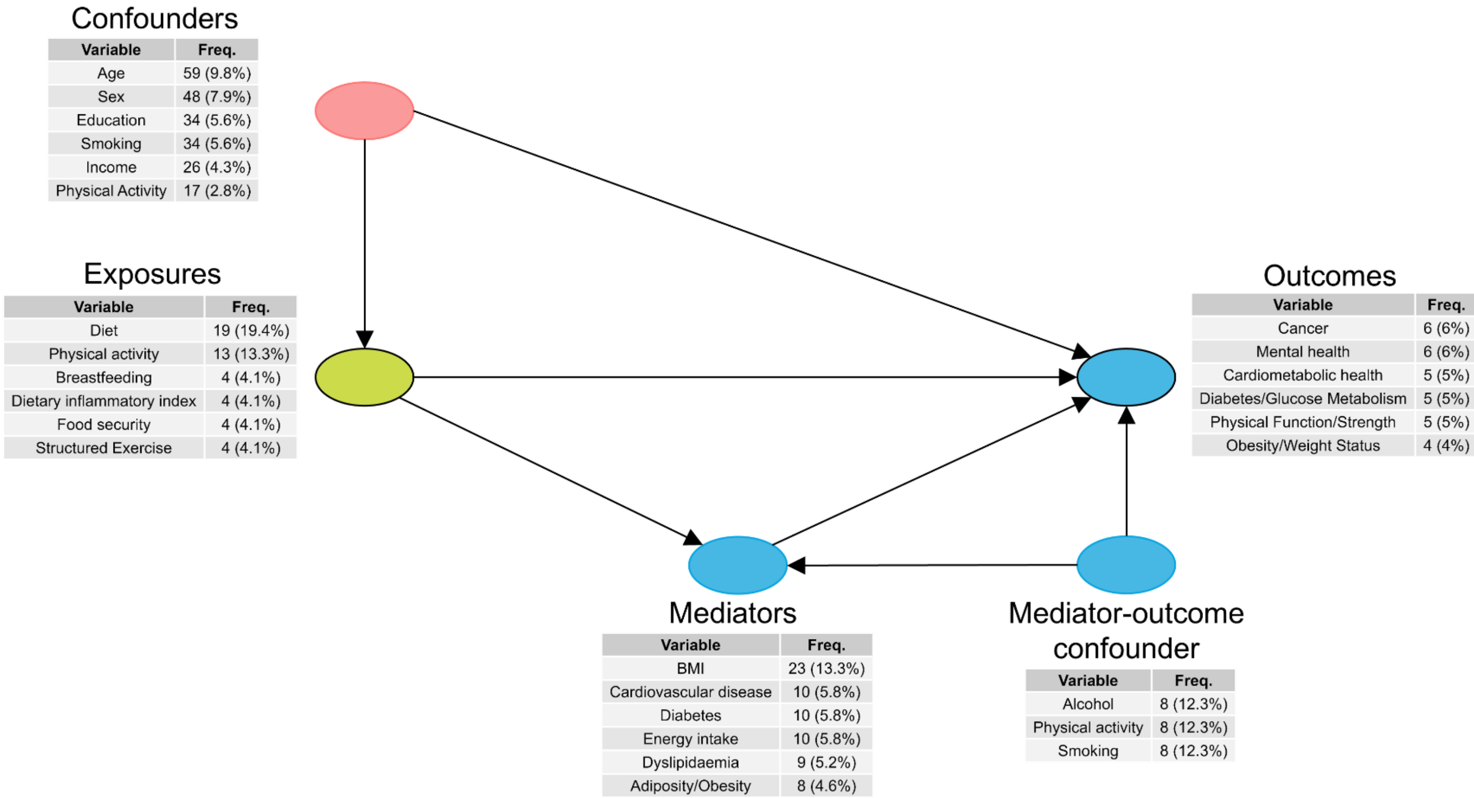


Figure 4. Visual representation of the most frequently reported variables across extracted DAGs.
Visual presents the common variables according to node type. Variables reported four or more times were selected to facilitate visualization.

Discussion

Summary of findings

Our scoping review identified a marked increase in the use of DAGs in physical activity and nutrition research in recent years. A major shortcoming in the DAG-building process was insufficient reporting of the methods used to construct DAGs. We also identified other issues of concern, such as studies not presenting their DAG, despite reporting using one, or presenting DAGs that were not acyclic. Regarding common structures found in DAGs, we identified limited consideration of key variables, such as mediators and mediator-outcome confounders. DAGs were more common in nutrition-focused than in physical activity-related research and were typically used in observational designs, namely, cross-sectional or cohort studies. We consider these findings and provide recommendations to enhance and expand the use of DAGs in these fields.

The DAG-building process

An important driver of DAG usefulness is the processes by which they are built, including how researchers determine which nodes and arcs to include. We identified that most studies (55%) provided no explanation regarding this methodological step. When explanations were available, authors typically stated that decisions were informed by a literature review, with varying degrees of transparency on how the literature was used. A prototypical example of insufficient reporting is the statement: "*a directed acyclic graph was developed (...) based on expert knowledge and the literature*". This provides little context on the source or role of expert input, how it informed the DAG structure, or how the literature was used to support specific arcs. In contrast, examples of greater methodological transparency included the provision of decision logs (20,24–26) linking individual causal arcs to specific references or theoretical assumptions.

Three studies (17,23,24) reported employing the ESC-DAGs method (13). This approach follows systematic review principles and begins with a fully saturated DAG, meaning all possible arcs are initially included. A step-by-step evaluation process (including counterfactual thought experiments and theory) is then used to remove implausible arcs. While methodologically rigorous, this process can be technically demanding. For example, one publication (27) reported that the DAGitty software occasionally failed when handling DAGs with an extensive number of nodes and arcs. This underscores the potential utility of computational approaches for data-driven causal structure learning, such as the PC-algorithm, a tool designed for building high-dimensional DAGs (28,29). Recently, however, a simulation study showed that while these tools have the potential to arrive at similar adjustment sets and estimated effects when compared to expert-based assessments, they also frequently produce inappropriate adjustment sets (30). As such, and although promising, causal discovery algorithms should be used with caution and may require manual refinement by topic-specific experts.

Given that one of the main strengths of DAGs is making analysts' assumptions explicit, and our current findings of insufficient information on the DAG-building process in most papers, we recommend that comprehensive guides, such as ESC-DAGs (31) and the framework proposed by Poppe et al. (32), be followed by researchers in physical activity and nutrition research in order to improve both methodological rigor and transparency.

Common structures of observed DAGs

The most frequently represented nodes identified in this analysis were confounders, followed by a smaller number of mediators, whereas competing exposures, instrumental variables and mediator-outcome confounders were relatively uncommon. This pattern suggests that DAGs were primarily used to identify confounders rather than to depict more complete causal structures. While identifying confounders is crucial, focusing solely on them risk overlooking key variables such as mediators and colliders that are essential for understanding the full set of causal pathways between exposure and outcome. Mediators are critical variables that account for part of the effect of an exposure on an outcome. When estimating the total effect, mediators should not be adjusted for, as they transmit part of the exposure's effect. In contrast, if the goal is to estimate the direct effect, mediators must be adjusted for to isolate on the exposure's direct impact. For example, the effect of physical activity on cardiovascular mortality may be partially mediated by changes in blood pressure. Whether or not to adjust for blood pressure depends on the analytic objective. If the aim is to estimate the total effect of physical activity, then blood pressure should not be adjusted for, as it lies on the causal pathway. In contrast, estimating the direct effect requires adjusting for blood pressure to isolate the portion of the effect that does not operate through this mediator.

An additional level of structure that was rarely included in studies was the potential for collider bias when adjusting for mediators. If another variable acts as a common cause of both the mediator and the outcome, it becomes a mediator-outcome confounder. Since the mediator is necessarily caused by the exposure, this opens a backdoor path and generates bias between exposure and outcome. If mediator-outcome confounders are correctly identified, however, adjusting for these variables will avoid the introduction of collider bias, even when estimating direct effects. Authors are therefore recommended to contemplate the role of mediators in their causal questions, while also interrogating the presence of mediator-outcome confounders to avoid potential collider bias.

We noted that demographic variables such as age, sex or gender, education and income were commonly identified as confounders. Additionally, common mediators included BMI, chronic diseases and energy intake. This is reasonable, as the effects of diet and/or physical activity on common outcomes (such as cancer, mental health, strength) might be due, in part, to indirect effects mediated by changes in body composition, status of chronic diseases, and energy intake. Similarly, the most common mediator-outcome confounders included alcohol intake, amount of physical activity, and smoking. Although we did not formally assess whether authors correctly applied the concepts of confounders and mediators, we observed instances where mediators (as depicted in the DAG) were inadvertently treated as confounders. In this case, wrongfully adjusting for mediators may lead to an underestimation of the total effect of the exposure. Similarly, certain DAGs aggregated variables into broad categories (*e.g.,* "demographics") rather than listing specific variables separately. This can be misleading unless all grouped variables (*e.g.,* "age", "ethnicity", "income") share the same causal structure.

Energy intake is a common and nuanced example of how researchers accommodate mediator variables in their analyses. It is standard practice to adjust for total energy intake when estimating the effect of a specific macronutrient, since carbohydrates, fats and proteins are components of total energy intake. However, doing so can introduce collider bias (33). This highlights the importance of carefully considering the causal relationships between variables when constructing a DAG, as traditional adjustment methods for energy intake can distort causal estimates (33). To mitigate this, Tomova et al. (33) propose that all macronutrients are included in the causal model. This also enables "substitution analysis", that is, estimating the effect of increasing one nutrient while reducing another (34,35). Not only does this avoid bias, but it also aligns with how nutrition RCTs and real-world dietary recommendations are structured: to increase one nutrient, another must typically be reduced to maintain energy balance. This logic is not limited to total energy intake, but can also be applied to any set of mutually constrained dietary exposures, such as saturated and unsaturated fat, or ultra-processed and minimally processed foods.

Common issues of observed DAGs

We identified several recurring issues across the studies analyzed in our review. First, there were issues with transparency and reproducibility, with approximately one-fifth of studies failing to provide their DAGs. Second, 9% of graphs contained cycles, violating the core assumption of acyclicity, which states that a variable cannot cause itself (36). Authors might encounter difficulty in removing loops from their DAG structures if they believe that certain variables influence each other reciprocally. For example, caloric intake increases body mass, but an increased body mass could, in turn, also impact caloric intake, as larger individuals eat more to maintain their current body mass. In this case, identifying the causal sequence of these effects often hinges on establishing the correct temporal ordering of variables. Constructing the DAG from left to right, with the x-axis representing the passage of time may aid in resolving cyclicality (6,32).

If a study explores various parameters, each with its own underlying assumptions, then it is important to construct dedicated DAGs that are specific to each outcome of interest. Despite this, very few studies presented multiple DAGs, even when using multiple statistical models with varying assumptions. This limits the capacity of the DAG to inform the most appropriate statistical analysis, considering that each outcome will have its own causal structure and specific confounding and mediating variables. In these cases, it stands to reason that the different assumptions under each model should be accompanied by altered DAGs illustrating each scenario. For instance, a study that aims to quantify the impact of physical activity levels on the separate outcomes of insulin sensitivity and quality of life would necessitate two different DAGs illustrating each scenario, given that the causal structure surrounding each outcome is likely to be different. Also, given that the true causal structure is unknown, creating separate DAGs for each model—or applying sensitivity analyses under different causal assumptions—could strengthen causal inferences (13,32).

Expanding the application of DAGs

This review found that most studies used DAGs in observational studies, whereas they were less commonly employed in other designs such as RCTs. Certainly, DAGs have clear application in observational designs; however, it is worth considering how they might also be used to visualize and mitigate bias in other designs. For example, Lee et al. (37) proposed a DAG-based approach for experimental studies that provides insight into potential strategies to handle missing data by incorporating the hypothesized mechanisms underlying missingness. Additionally, Bulbulia (38) recently described how DAGs can be used to visualize and mitigate biases that commonly arise in RCTs. One example relates to per-protocol analyses—when an RCT analyzes individuals according to their adherence to the intervention, rather than including all participants regardless of adherence. In this scenario, bias may arise if a confounder influences both adherence to the intervention and the outcomes. An unbiased per-protocol effect can be recovered if the analyst can identify and condition on the confounder in question, or a proxy thereof (38).

One study included in our review, Evanchuk et al (39) included an effect modifier in the visual representation of their DAG. Modifiers are understood as variables that alter the magnitude and potentially the directionality (*e.g.* positive/negative effect) of the causal effect between exposure and outcome. This is in contrast with mediators, which carry part of the effect of the exposure, but don't change the total effect of the exposure. In this case, authors were interested in whether maternal iron biomarkers differently affected newborn birthweight according to their sex. While recent work has suggested methods for integrating effect modifiers into DAGs (40), this remains a limitation in widely used tools such as DAGitty, which currently lack functionality to represent these interaction effects.

Conclusion

The use of DAGs in physical activity and nutrition research-related research has grown in recent years. While DAGs are valuable for identifying adjustment sets, their practical utility depends on accurately representing the underlying causal structure. Many of the studies we reviewed did not sufficiently describe or employ a systematic DAG-building process, often omitted critical nodes (*e.g.,* mediators, mediator–outcome confounders), and, in some cases, did not make their DAGs publicly available.

To address these issues, we recommend that researchers: 1) Clearly document their DAG-building process, for instance with a decision log, and to follow established guidelines and structured frameworks, such as ESC-DAGs (31) or the approaches outlined by Poppe et al (32); 2) Incorporate temporal information by arranging variables chronologically to reflect causal order and to avoid cyclical paths; 3) If multiple models or outcomes are of interest, use separate DAGs to clearly depict the potential causal structure and underlying assumptions of each one; 4) Consider all relevant variables, which in addition to confounders, includes mediators, mediator-outcome confounders, colliders, and effect modifiers. These latter variables are particularly relevant when estimating total versus direct effects; and 5) Expand DAG use to identify potential sources in bias in RCTs, including those related to adherence, missing data, and selection issues in randomized trials. Following these recommendations may enhance the transparency, reproducibility, and accuracy of DAG-based causal inference in physical activity and nutrition research.

Acknowledgements

G. Esteves, P. Swinton and E. Dolan conceptualized the idea for the review. P. Swinton and L. Santos assisted with methodology for the review. L. Santos assisted with conceptualizing the search strategy. G. Esteves wrote the initial draft of the manuscript. G. Esteves, P. Swinton and J. Shim screened articles for eligibility. G. Esteves and J. Slaton extracted the data. G. Esteves analysed the resulting data and created the visualizations and summaries. All authors contributed by reviewing, editing, and writing the final version of the manuscript.

Funding

G. Esteves, E. Dolan and L. Santos are supported by research grants from the São Paulo Research Foundation (FAPESP, grant numbers 2020/07860-9, 2019/05616-6, 2019/26899-6, 2023/08433-5). No specific funding was provided for this investigation.

Conflicts of interest

The authors declare no conflict of interest.

References

1. Pearl J. Causal Inference in the Health Sciences: A Conceptual Introduction. Health Serv Outcomes Res Methodol. 2001;2(3/4):189–220.
2. Hernán MA, Hsu J, Healy B. A Second Chance to Get Causal Inference Right: A Classification of Data Science Tasks. CHANCE. 2019 Jan 2;32(1):42–9.
3. Grimes DA, Schulz KF. Bias and causal associations in observational research. The Lancet. 2002 Jan;359(9302):248–52.
4. Hernán MA, Hernández-Díaz S, Robins JM. A Structural Approach to Selection Bias. Epidemiology. 2004 Sep;15(5):615–25.
5. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology. 1999 Jan;10(1):37–48.
6. Tennant PWG, Murray EJ, Arnold KF, Berrie L, Fox MP, Gadd SC, et al. Use of directed acyclic graphs (DAGs) to identify confounders in applied health research: review and recommendations. Int J Epidemiol. 2021 May 17;50(2):620–32.
7. Knüppel S, Stang A. DAG program: identifying minimal sufficient adjustment sets. Epidemiology. 2010 Jan;21(1):159.
8. Hernan MA, Cole SR. Invited Commentary: Causal Diagrams and Measurement Bias. Am J Epidemiol. 2009 Oct 15;170(8):959–62.
9. Peters MM, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews. In: JBI Manual for Evidence Synthesis. JBI; 2020.
10. Clark JM, Sanders S, Carter M, Honeyman D, Cleo G, Auld Y, et al. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. Journal of the Medical Library Association. 2020 Apr 1;108(2).
11. Tricco AC, Lillie E, Zarin W, O’Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med. 2018 Oct 2;169(7):467–73.
12. Textor J, Hardt J, Knüppel S. DAGitty. Epidemiology. 2011 Sep;22(5):745.

13. Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. Int J Epidemiol. 2020 Feb 1;49(1):322–9.

14. Tennant PW, Arnold K, Berrie L, Ellison GT, Gilthorpe MS. ADVANCED MODELLING STRATEGIES Challenges and pitfalls in robust causal inference with observational data [Internet]. 2017. Available from: http://lida.leeds.ac.uk/

15. Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, et al. Assessment of Bidirectional Relationships Between Physical Activity and Depression Among Adults. JAMA Psychiatry. 2019 Apr 1;76(4):399.

16. Pearl J. Causal Diagrams for Empirical Research. Biometrika. 1995 Dec;82(4):669.

17. Campbell T, Cullen B. Estimating the effect of physical activity on cognitive function within the UK Biobank cohort. Int J Epidemiol. 2023 Oct 5;52(5):1592–611.

18. Foraita R, Witte J, Börnhorst C, Gwozdz W, Pala V, Lissner L, et al. A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents. Sci Rep. 2024 Mar 21;14(1):6822.

19. Peng Z, Apfelbacher C, Brandstetter S, Eils R, Kabesch M, Lehmann I, et al. Directed acyclic graph for epidemiological studies in childhood food allergy: Construction, user's guide, and application. Allergy. 2024 Aug 17;79(8):2051–64.

20. Schwartz MH, Steele KM, Ries AJ, Georgiadis AG, MacWilliams BA. A model for understanding the causes and consequences of walking impairments. PLoS One. 2022 Dec 28;17(12):e0270731.

21. Horgan BG, Drew MK, Halson SL, Piromalli LE, Drinkwater EJ, Chapman DW, et al. Impaired recovery is associated with increased injury and illness: A retrospective study of 536 female netball athletes. Scand J Med Sci Sports. 2021 Mar 18;31(3):691–701.

22. Amaral YN di V do, Marano D, Filha MMT, Moreira MEL. Pre-gestational overweight and polyunsaturated fatty acids in human milk: theoretical causality model. Cien Saude Colet. 2024;29(2).

23. Tang D, Xiao X, Chen L, Kangzhu Y, Deng W, Basang, et al. Association of dietary patterns with obesity and metabolically healthy obesity phenotype in Chinese population: a cross-sectional analysis of China Multi-Ethnic Cohort Study. British Journal of Nutrition. 2022 Dec 14;128(11):2230–40.

24. Xiao X, Qin Z, Lv X, Dai Y, Ciren Z, Yangla Y, et al. Dietary patterns and cardiometabolic risks in diverse less-developed ethnic minority regions: results from the China Multi-Ethnic Cohort (CMEC) Study. Lancet Reg Health West Pac. 2021 Oct;15:100252.

25. Preußel K, Milde-Busch A, Schmich P, Wetzstein M, Stark K, Werber D. Risk Factors for Sporadic Non-Pregnancy Associated Listeriosis in Germany—Immunocompromised Patients and Frequently Consumed Ready-To-Eat Products. PLoS One. 2015 Nov 23;10(11):e0142986.

26. Kwok WS, Khalatbari-Soltani S, Dolja-Gore X, Byles J, Oliveira JS, Pinheiro MB, et al. Differences in Falls and Physical Activity in Older Women From Two Generations. J Gerontol A Biol Sci Med Sci. 2024 Apr 1;79(4).

27. Campbell T, Ferguson K, Whyte J, Cullen B. Constructing a graphical model of the relationship between physical activity and cognitive function based on a systematic review of prospective evidence. 2021.

28. Foraita R, Witte J, Börnhorst C, Gwozdz W, Pala V, Lissner L, et al. A longitudinal causal graph analysis investigating modifiable risk factors and obesity in a European cohort of children and adolescents. Sci Rep. 2024 Mar 21;14(1):6822.

29. Kalisch M, Ch BME. Estimating High-Dimensional Directed Acyclic Graphs with the PC-Algorithm Peter B ¨ uhlmann. Vol. 8, Journal of Machine Learning Research. 2007.

30. Gururaghavendran R, Murray EJ. Can algorithms Replace Expert Knowledge for Causal Inference? A Case Study on Novice Use of Causal Discovery. Am J Epidemiol. 2024 Aug 31;

31. Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, et al. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. Int J Epidemiol. 2020 Feb 1;49(1):322–9.

32. Poppe L, Steen J, Loh WW, Crombez G, De Block F, Jacobs N, et al. How to develop causal directed acyclic graphs for observational health research: a scoping review. Health Psychol Rev. 2024 Sep 27;1–21.

33. Tomova GD, Arnold KF, Gilthorpe MS, Tennant PWG. Adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr [Internet]. 2022 Jan 11;115(1):189–98. Available from: https://academic.oup.com/ajcn/article/115/1/189/6329041

34. Tomova GD, Gilthorpe MS, Tennant PW. Theory and performance of substitution models for estimating relative causal effects in nutritional epidemiology. Am J Clin Nutr. 2022 Nov;116(5):1379–88.

35. Ibsen DB, Dahm CC. Food substitutions revisited. Am J Clin Nutr. 2022 Nov;116(5):1195–8.

36. PEARL J. Causal diagrams for empirical research. Biometrika. 1995;82(4):669–88.

37. Lee KJ, Carlin JB, Simpson JA, Moreno-Betancur M. Assumptions and analysis planning in studies with missing data in multiple variables: moving beyond the MCAR/MAR/MNAR classification. Int J Epidemiol. 2023 Aug 2;52(4):1268–75.

38. Bulbulia JA. Methods in causal inference. Part 4: confounding in experiments. Evol Hum Sci. 2024 Sep 27;6:e43.

39. Evanchuk JL, Kozyrskyj A, Hanas N, Goruk S, Vaghef-Mehrabani E, Archundia-Herrera CM, et al. Maternal Iron Status Is Dynamic Throughout Pregnancy and Might Predict Birth Outcomes in a Sex Dependent Manner: Results from the Alberta Pregnancy Outcomes and Nutrition (APrON) Cohort Study. J Nutr. 2023 Sep;153(9):2585–97.

40. Nilsson A, Bonander C, Strömberg U, Björk J. A directed acyclic graph for interactions. Int J Epidemiol. 2021 May 17;50(2):613–9.