Supplementary File 1. DAG Scoping Review searches (conducted on March 1st, 2024)

1. Medline/PubMed

#	Search String	Hits
1	exercise[mesh] OR "exercise therapy"[mesh] OR "physical activity" OR "exercise training" OR "aerobic training" OR "aerobic exercis*" OR "strength training" OR "strength exercise" OR "resistance training" OR "muscle stretching exercis*" OR "physical conditioning, human" OR "high-intensity interval training" OR HIIT OR "endurance training" OR running[mesh] OR walking[mesh] OR swimming[mesh] OR "isometric exercis*" OR "physical exercis*" OR "sports"[mesh] OR "cardiorespiratory fitness" OR "physical fitness"	537,700
2	Diet[mesh] OR "Diet*" OR "Eating"[mesh] OR "Eating" OR "Dietary Patterns"[mesh] OR "Dietary Patterns" OR "Nutrition" OR "Nutrient*" OR "Nutrition Therapy"[mesh] OR Nutritional Physiological Phenomena[mesh] OR "Nutritional Requirements"[mesh] OR "Nutritional Requirements" OR "Sports Nutritional Physiological Phenomena"[mesh]	1,814,712
3	#1 OR #2	2,245,967
4	"graphical model theory"[All Fields] OR "directed acyclic graph*"[All Fields] OR "causal diagram*"[All Fields] OR "causal graph*"[All Fields] OR "causal DAG"[All Fields]	1,755
5	#3 AND #4	142

2. Embase (Elsevier)

#	Translated (Polyglot) Search String	Hits
1	((exercise/exp OR 'exercise therapy'/exp OR 'physical activity' OR 'exercise training' OR 'aerobic training' OR 'aerobic exercis*' OR 'strength training' OR 'strength exercise' OR 'resistance training' OR 'muscle stretching exercis*' OR 'physical conditioning, human' OR	418

	'high-intensity interval training' OR HIIT OR 'endurance training' OR running/exp OR walking/exp OR swimming/exp OR 'isometric exercis*' OR 'physical exercis*' OR sports/exp OR 'cardiorespiratory fitness' OR 'physical fitness') OR (Diet/exp OR Diet* OR Eating/exp OR Eating OR 'Dietary Patterns'/exp OR 'Dietary Patterns' OR Nutrition OR Nutrient* OR 'Nutrition Therapy'/exp OR 'Nutritional Physiological Phenomena'/exp OR 'Nutritional Requirements'/exp OR 'Nutritional Requirements' OR 'Sports Nutritional Physiological Phenomena'/exp)) AND ('graphical model theory' OR 'directed acyclic graph*' OR 'causal diagram*' OR 'causal graph*' OR 'causal DAG')	

3. Cochrane Central

#	Translated (Polyglot) Search String	Hits
1	(([mh exercise] OR [mh "exercise therapy"] OR "physical activity" OR "exercise training" OR "aerobic training" OR ("aerobic" NEXT exercis*) OR "strength training" OR "strength exercise" OR "resistance training" OR ("muscle stretching" NEXT exercis*) OR "physical conditioning, human" OR "high-intensity interval training" OR HIIT OR "endurance training" OR [mh running] OR [mh walking] OR [mh swimming] OR ("isometric" NEXT exercis*) OR ("physical" NEXT exercis*) OR [mh sports] OR "cardiorespiratory fitness" OR "physical fitness") OR ([mh Diet] OR Diet* OR [mh Eating] OR Eating OR [mh "Dietary Patterns"] OR "Dietary Patterns" OR Nutrition OR Nutrient* OR [mh "Nutrition Therapy"] OR [mh "Nutritional Physiological Phenomena"] OR [mh "Nutritional Requirements"] OR "Nutritional Requirements" OR [mh "Sports Nutritional Physiological Phenomena"])) AND ("graphical model theory" OR ("directed acyclic" NEXT graph*) OR ("causal" NEXT diagram*) OR ("causal" NEXT graph*) OR "causal DAG")	4

4. SPORTDiscus

#	Translated (Polyglot) Search String	Hits
1	((DE "exercise" OR DE "exercise therapy" OR "physical activity" OR "exercise training" OR "aerobic training" OR "aerobic exercis*" OR "strength training" OR "strength exercise" OR "resistance training" OR "muscle stretching exercis*" OR "physical conditioning, human" OR "high-intensity interval training" OR "HIIT" OR "endurance training" OR DE "running" OR DE "walking" OR DE "swimming" OR "isometric exercis*" OR "physical exercis*" OR DE "sports" OR "cardiorespiratory fitness" OR "physical fitness")	84

	OR (DE "Diet" OR "Diet*" OR DE "Eating" OR "Eating" OR DE "Dietary Patterns" OR "Dietary Patterns" OR "Nutrition" OR "Nutrient*" OR DE "Nutrition Therapy" OR DE "Nutritional Physiological Phenomena" OR DE "Nutritional Requirements" OR "Nutritional Requirements" OR DE "Sports Nutritional Physiological Phenomena")) AND (TX "graphical model theory" OR TX "directed acyclic graph*" OR TX "causal diagram*" OR TX "causal graph*" OR TX "causal DAG")	

Supplementary File 2. Codebook detailing the description of which extracted variable and how they were ascertained

Column	Heading	Description
A	ID	Unique identifier for each study entry.
B	DOI	Digital Object Identifier (DOI) linking to the study publication.
C	Author	Last name of the first author.
D	Year	Publication year of the study.
E	Title	Title of the study.
F	Country	Country associated with the paper. When coauthors were affiliated with different countries, the affiliation of the first author was used.
G	Journal	Name of the journal where the study was published.
H	Aims/Purpose	Brief description of the study's objectives or research questions, usually copied and pasted from their own statements.
I	Area	Broad field or discipline within each exposure type (nutrition or exercise science). Options were “Biomechanics” (studies involving movement impairments and their determinants); Exercise epidemiology (studies assessing the relevance of physical activity levels on disease incidence, risk or mortality in large populations); Exercise physiology (studies assessing the role of physical activity/exercise-related exposures on physiological responses, such as inflammation, bone turn over, etc.); Sports medicine (studies assessing the role of physical activity/exercise-related exposures on sports-related outcomes, such as athlete injury and recovery); Strength and conditioning (studies assessing the role of physical activity/exercise-related exposures on strength and condition-related

Column	Heading	Description
		variables, such as handgrip or bench press strength); Nutritional epidemiology (studies assessing the role of nutritional exposures on disease incidence, risk or mortality in large populations); and Clinical nutrition (studies assessing the role of nutritional exposures on clinical outcomes of individuals, such as allergies).
J	Study design	Study design of the paper. Options were "Case-control" (observational study that match individuals with a condition of interest with similar controls), "Causal model creation study" (studies aimed at creating an overall causal model for a given problem, without necessarily conducting data analysis), Longitudinal cohort (observational studies with a longitudinal design); Cross-sectional (observational study with a cross-sectional design); Mendelian randomization (observational studies that use mendelian randomization techniques, i.e., using gene variation to determine causal effects); Randomized controlled trial (trials that employ randomization and control/placebo groups to assess the effect of interventions); Systematic review and/or meta-analysis (studies that conduct systematic review and/or meta-analytic combination of the included studies).
K	Data source	Source of the data (e.g., NHANES, UK Biobank, Author's own data, etc.).
L	Data size (sample)	Total number of participants in the study.
M	DAG availability	Indicates whether the DAG was available in the paper, supplementary material or any other online repository (Yes/No).
N	Reason for DAG (text)	Verbatim explanation given by authors of why a DAG was employed.
O	DAG software	Software/tool used to create the DAG (e.g., DAGitty, ggdag).
P	DAG develop method statement	Verbatim explanation given by authors on how the DAG was constructed.

Column	Heading	Description
Q	DAG development method	Categorization of the DAG development statement (full text obtained from previous column). Options were “Not reported” (when authors simply did not state any method for building the DAG); “Literature-based” (when authors claimed using some type of literature assessment to construct the DAG); “Expert-based” (when authors claimed using their own or some other experts’ knowledge to construct the DAG); “Literature and expert-based” (when authors claimed using a combination of both previous methods); “Delphi-consensus” (when authors claimed using a standardized Delphi panel method to create a consensus and build the DAG); “Causal discovery algorithm” (when authors used and automatized algorithm to discover/build the DAG, e.g., the PC-algorithm); “Other” (some other method).
R	Number DAGs	Total number of DAGs included in the study. If more than one DAG was available per paper, each separate DAG was extracted in a new row, and the studies name, ID and title were repeated.
S	Number nodes	Number of nodes (variables) in the DAG.
T	Outcome	Primary outcome variable in the study (e.g., Cancer, Cardiovascular risk).
U	Exposure	Primary exposure variable (e.g., Ultra-processed food intake, Physical activity levels).
V	Exposure category	Category of exposure (Nutrition or Exercise science).
W	Exposure measure	Method or tool used to measure the exposure (e.g., food frequency questionnaire).
X	Number causal paths (exposure to outcome)	Count of causal paths from exposure to outcome in the DAG.
Y	Direct	Indicates if a direct causal path exists (Yes/No).
Z	Shortest path	Length of the shortest causal path available (number of edges).

Column	Heading	Description
AA	Longest path	Length of the longest causal path available (number of edges).
AB	Number mediators	Count of mediator variables in the DAG. Mediators are variable that are found in the causal path between the exposure and the outcome, i.e., are caused by the exposure, and cause the outcome.
AC	Identified mediators	Names of mediator variables (if any). Multiple entries per cell were allowed and were separated by a comma.
AD	Number of mediator-outcome confounders	Count of confounders between mediators and outcome. Mediator-outcome confounders are variables point to both the mediator and the outcome, i.e., cause both the mediator and the outcome. Authors did not require to specifically mention the presence of mediator-outcome confounders; rather, these were noted/extracted if they were visually identified in the DAG by the researchers. This approach was followed for all variable types.
AE	Identified mediator-outcome confounders	Names of mediator-outcome confounders (if any). Multiple entries per cell were allowed and were separated by a comma. Multiple entries per cell were allowed and were separated by a comma.
AF	Estimating total or direct effect	Whether the study estimated total, direct, or both effects. This was ascertained by: explicit mention of it by authors; or, when this was not available, estimation of total effects was assumed when no mediators were present, or when they were presented but were not conditioned when; direct effects were assumed when mediators were present and conditioned on; or both, when both approaches were utilized in the same paper.
AG	Number confounders	Total count of confounder variables in the DAG. Confounders were considered as variables that cause both the exposure and the outcome, i.e., point to both the exposure and the outcome.
AH	Identified confounders	Names of confounder variables. Multiple entries per cell were allowed and were separated by a comma.

Column	Heading	Description
AI	Number of competing exposures	Count of competing exposures considered. Competing exposures were considered as variables that cause the outcome, but are not linked to the exposure in any way.
AJ	Identified competing exposures	Names of competing exposures (if any). Multiple entries per cell were allowed and were separated by a comma.
AK	Number instrumental variables	Count of instrumental variables. Instrumental variables were considered as variables that cause the exposure, but are totally independent from the outcome.
AL	Identified instrumental variables	Names of instrumental variables (if any). Multiple entries per cell were allowed and were separated by a comma.
AM	Number of edges	Total edges (arrows) in the DAG.
AN	Is DAG acyclical	Indicates if the DAG contains no cycles or loops (Yes/No).
AO	Statistical model	Model used for the analysis (e.g., generalized linear model (GLM) linear model; GLM logistic model; cox regression). Multiple entries per cell were allowed and were separated by a comma.
AP	Treatment-outcome relationship	Description of the relationship between exposure and outcome. Options were “categorical regression”, when the exposure was categorized into groups (e.g., quartiles of intake of ultra-processed foods); “continuous linear” when the exposure was continuous and a linear trend was estimated; “continuous with rcs” when the exposure was continuous and the model used a restricted cubic spline approach; “continuous polynomial” when the exposure was continuous and the model used a polynomial approach; “with interaction” was also noted when either a linear, polynomial or RCS fit was had an interaction with another variable. Multiple entries per cell were allowed and were separated by a comma.
AQ	Adjustment method	Method for adjusting confounders, such as covariate adjustment in multivariable regression; Inverse

Column	Heading	Description
		probability weighting; Parametric g-formula; No statistical analysis performed (in papers that included DAGs but no analysis).
AR	Number of categories (if categorical)	For categorical exposure variables, the number of levels were noted (e.g., if binary, 2 levels, or if using quartiles, 4 levels, etc.). Multiple entries per cell were allowed and were separated by a comma.
AS	Category labels	Labels for categorical variable levels (when applicable).
AT	Adjustment set reported	Indicates if the adjustment set was explicitly reported somewhere in the text (Yes/No). This was ascertained by inspecting the manuscript either for explicit mentions of “adjustment sets”, or when authors reported variables included in statistical models in the methods or results section, in table or figure legends, etc.
AU	Causal or associational language used	Whether the language surrounding statements in the paper were mostly causal or associational. To ascertain this, the reporting of results and conclusion from the papers were evaluated. Statements such as “The exposure was associated with a change in the outcome” were classified as associational language, whereas “The exposure led to a change the outcome” or “The exposure had an effect on the outcome” were classified as causal language. In papers wherein both types of statements where present, the most prevalent was used.

Supplementary Table S1. Recoding of extracted DAG variables

Original extracted variable name	**Recoded as**
Confounders	
"age", "age group", "baseline age"	Age
"sex", "gender"	Sex
"education", "educational level", "schooling", "schooling level", "graduation level", "education level", "degree of education", "educational attainment"	Education
"smoking", "smoking status", "smoking habit", "smoking history", "regular smoking", "smoke", "cigarette smoking", "waterpipe smoking", "tabacco exposure"	Smoking
"alcohol", "alcohol intake", "alcohol consumption"	Alcohol consumption
"physical activity", "leisure-time physical activity", "occupational pa", "physical activity level", "exercise", "leisure-time physical activity"	Physical activity
"bmi", "pre-pregnancy bmi", "prepregnancy bmi"	BMI
"body weight", "birthweight", "birth weight"	Body weight
"adiposity", "obesity"	Adiposity/Obesity
"socioeconomic status (ses)", "ses", "economic class", "social class", "socioeconomic status", "socioeconomic level", "index of multiple deprivation", "townsend deprivation index"	Socioeconomic status
"income", "household income", "family income", "wealth assets index", "wealth index", "average household income", "brazilian economic classification"	Income
"diet", "food consumption", "diet pattern", "household nutritional habits", "pre-pregnancy dietary intake", "overall diet quality", "food consumption pattern"	Diet
"mental health", "depressive symptoms", "anxiety", "stress", "anxiety symptoms", "mood_t1"	Mental health
"comorbidities", "number of morbidities", "history of diabetes", "diabetes", "heart diseases", "stroke"	Comorbidities/Chronic diseases
"ethnicity", "skin color"	Ethnicity
Outcomes	
"depressive symptoms", "depression", "mental health problem", "psychological well-being", "phychological symptoms"	Mental health
"obesity phenotype", "overweight_obesity", "obesity and excess body fat", "child's weight status"	Obesity/Weight Status
"type 2 diabetes", "diabetes", "insulin resistance", glucose and lipid biomarkers"	Diabetes/Glucose Metabolism
"cardiometabolic risk", "metabolic syndrome", "hypertension", "blood pressure", "ischemic stroke"	Cardiometabolic health
"colorectal cancer", "pancreatic cancer", "ovarian cancer", "acute leukemia"	Cancer

"infant development", "offspring cognition", "child language and learning"	Child development
"handgrip strength", "muscle strength index", "physical function", "grip strength decline"	Physical Function/Strength
"gestational hypertension", "birth defects", "adverse birth outcomes", "small-for-gestational-age birth"	Pregnancy/Birth Outcomes
"dietary pattern adherence", "healthfulness of food choice", "prudent diet score"	Diet
"mortality", "mortality risk"	Mortality
Exposures	
"dietary intake", "diet", "dietary pattern", "dietary patterns", "food consumption", "consumption of food", "changes of adherance to dietary patterns"	Diet
"dairy intake", "milk and milk products"	Dairy
"coffee consumption", "caffeine"	Coffee/Caffeine
"added sugar intake", "fructose intake", "honey consumption"	Sugar
"ultra-processed food consumption", "ultra-processed foods", ultra-processed food intake"	Ultra-processed foods
"exclusive breastfeeding", "breastfeeding duration"	Breastfeeding
"physical activity", "leisure-time physical activity", "physical activity_0", "steps"	Physical activity
"exercise therapy vs no-exercise control", "running", "6-meter timed hop test"	Structured Exercise
"maternal exercise in pregnancy", "exercise during pregnancy"	Pregnancy-specific exercise
"muscle mass", "sarcopenia"	Muscle mass/Sarcopenia
"food security", "nutrition security", "food insecurity"	Food security
Mediators	
"age", "age group", "baseline age", "age at first full pregnancy", "age at menarch"	Age
"bmi", "body mass index", "current bmi", "bmi pre-pregnancy", "bmi_3", "bmi_4", "bmi_5", "bmi_6", "infant bmi"	BMI
"weight", "weight of the child at birth"	Body weight
"lean mass", "muscle mass"	Muscle mass
"fat mass", "excess weight", "obesity", "adiposity", "abdominal obesity"	Adiposity/Obesity
"energy intake", "total energy intake", "total calorie intake", "energy intake body fat", "total energy"	Energy intake
"diabetes", "diabetes_3", "diabetes_4", "diabetes_5", "diabetes_6", "diabetes_dr_or_meas"	Diabetes
"ldl", "hypercholesterolemia", "tg", "high_cholestoreol_dr_or_meas", "dyslipidemia", "tg/hdl-c ratio	Dyslipidaemia
"cvd_3", "cvd_4", "cvd_5", "cvd_6", "hypertension"	Cardiovascular disease
"physical activity", "physical activity_1", "running activity"	Physical activity
"healthy diet score", "diet quality"	Diet quality

"family income", "seifa"	Socioeconomic factors
Mediator-outcome confounder	
"alcohol", "alcohol consumption", "alcohol drinking", "alcohol intake", "alcohol use"	Alcohol
"tobacco", "tobacco use", "smoking"	Smoking
Instrumental variables	
"Education", "Educational level", "education level"	Education
"total energy intake", "Energy intake from diet"	Energy intake
Competing exposures	
"alcohol", "alcohol consumption", "alcohol drinking", "alcohol intake", "alcohol use"	Alcohol
"tobacco", "tobacco use", "smoking"	Smoking

Supplementary Table S2. Complete information on included studies characteristics

Characteristic	**Overall, N = 115**	**Exercise, N = 38**	**Nutrition, N = 77**
Data sample size	2,487 (920, 16,376)	1,703 (592, 11,641)	2,550 (986, 19,654)
Country			
Australia	15 (13%)	6 (16%)	9 (12%)
Brazil	19 (17%)	3 (7.9%)	16 (21%)
Canada	5 (4.3%)	2 (5.3%)	3 (3.9%)
China	18 (16%)	3 (7.9%)	15 (19%)
Denmark	7 (6.1%)	3 (7.9%)	4 (5.2%)
Ethiopia	2 (1.7%)	0 (0%)	2 (2.6%)
Finland	1 (0.9%)	1 (2.6%)	0 (0%)
France	3 (2.6%)	0 (0%)	3 (3.9%)
Germany	4 (3.5%)	1 (2.6%)	3 (3.9%)
India	1 (0.9%)	0 (0%)	1 (1.3%)
Iran	2 (1.7%)	1 (2.6%)	1 (1.3%)
Italy	1 (0.9%)	0 (0%)	1 (1.3%)
Japan	2 (1.7%)	1 (2.6%)	1 (1.3%)
Lebanon	1 (0.9%)	0 (0%)	1 (1.3%)
Mexico	1 (0.9%)	0 (0%)	1 (1.3%)
Netherlands	3 (2.6%)	2 (5.3%)	1 (1.3%)
Norway	9 (7.8%)	5 (13%)	4 (5.2%)
Peru	1 (0.9%)	0 (0%)	1 (1.3%)
Saudi Arabia	1 (0.9%)	1 (2.6%)	0 (0%)
Singapore	1 (0.9%)	0 (0%)	1 (1.3%)
South Africa	1 (0.9%)	1 (2.6%)	0 (0%)
Spain	2 (1.7%)	1 (2.6%)	1 (1.3%)
Sweden	1 (0.9%)	1 (2.6%)	0 (0%)
UK	4 (3.5%)	2 (5.3%)	2 (2.6%)

USA	10 (8.7%)	4 (11%)	6 (7.8%)
Journal			
Acta Orthopaedica	1 (0.9%)	1 (2.6%)	0 (0%)
ADC Fetal & Neonatal	1 (0.9%)	0 (0%)	1 (1.3%)
Age and Ageing	1 (0.9%)	1 (2.6%)	0 (0%)
Allergy	1 (0.9%)	0 (0%)	1 (1.3%)
American Journal of Clinical Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
Annals of Epidemiology	1 (0.9%)	1 (2.6%)	0 (0%)
Atherosclerosis	1 (0.9%)	1 (2.6%)	0 (0%)
BMC Public Health	3 (2.6%)	0 (0%)	3 (3.9%)
Brazilian Journal of Medical and Biological Research	1 (0.9%)	0 (0%)	1 (1.3%)
British Journal of Dermatology	1 (0.9%)	1 (2.6%)	0 (0%)
British Journal of Nutrition	3 (2.6%)	0 (0%)	3 (3.9%)
British Journal of Sports Medicine	2 (1.7%)	2 (5.3%)	0 (0%)
British Medical Journal	1 (0.9%)	0 (0%)	1 (1.3%)
Cadernos de Saúde Pública	2 (1.7%)	1 (2.6%)	1 (1.3%)
Cancer Epidemiology	2 (1.7%)	0 (0%)	2 (2.6%)
Ciência & Saúde Coletiva	1 (0.9%)	0 (0%)	1 (1.3%)
Clinical Nutrition ESPEN	1 (0.9%)	0 (0%)	1 (1.3%)
Clinical Oral Investigations	1 (0.9%)	0 (0%)	1 (1.3%)
Community Dentistry and Oral Epidemiology	1 (0.9%)	0 (0%)	1 (1.3%)
Current Developments in Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
Diabetes Research and Clinical Practice	1 (0.9%)	1 (2.6%)	0 (0%)
eBioMedicine	1 (0.9%)	1 (2.6%)	0 (0%)
European Journal of Cancer	1 (0.9%)	0 (0%)	1 (1.3%)
European Journal of Clinical Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
European Journal of Nutrition	22 (19%)	0 (0%)	22 (29%)
Frontiers in Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
Frontiers in Oncology	1 (0.9%)	0 (0%)	1 (1.3%)

Gastroenterology	1 (0.9%)	0 (0%)	1 (1.3%)
International Journal of Behavioral Medicine	2 (1.7%)	2 (5.3%)	0 (0%)
International Journal of Cancer	1 (0.9%)	0 (0%)	1 (1.3%)
International Journal of Epidemiology	1 (0.9%)	1 (2.6%)	0 (0%)
International Journal of Sport Nutrition and Exercise Metabolism	1 (0.9%)	1 (2.6%)	0 (0%)
International Journal of Women's Health	1 (0.9%)	1 (2.6%)	0 (0%)
JAMA Internal Medicine	1 (0.9%)	1 (2.6%)	0 (0%)
JAMA Network Open	2 (1.7%)	1 (2.6%)	1 (1.3%)
Journal of Affective Disorders	1 (0.9%)	0 (0%)	1 (1.3%)
Journal of Developmental Origins of Health and Disease	1 (0.9%)	0 (0%)	1 (1.3%)
Journal of Human Nutrition and Dietetics	1 (0.9%)	0 (0%)	1 (1.3%)
Journal of International Medical Research	2 (1.7%)	2 (5.3%)	0 (0%)
Journal of Periodontology	1 (0.9%)	0 (0%)	1 (1.3%)
Journal of Physical Activity and Health	2 (1.7%)	2 (5.3%)	0 (0%)
Journal of Physiotherapy	1 (0.9%)	1 (2.6%)	0 (0%)
Journal of Science and Medicine in Sport	1 (0.9%)	1 (2.6%)	0 (0%)
Journal of Sport and Health Science	1 (0.9%)	1 (2.6%)	0 (0%)
Journal of Sports Sciences	1 (0.9%)	0 (0%)	1 (1.3%)
Knee Surgery, Sports Traumatology, Arthroscopy	1 (0.9%)	1 (2.6%)	0 (0%)
Maternal & Child Nutrition	2 (1.7%)	0 (0%)	2 (2.6%)
Maturitas	1 (0.9%)	0 (0%)	1 (1.3%)
Nutrients	7 (6.1%)	1 (2.6%)	6 (7.8%)
Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
Nutrition and Cancer	1 (0.9%)	0 (0%)	1 (1.3%)
Nutrition Journal	1 (0.9%)	0 (0%)	1 (1.3%)
Nutrition Research	1 (0.9%)	0 (0%)	1 (1.3%)
Nutrition, Metabolism & Cardiovascular Diseases	1 (0.9%)	0 (0%)	1 (1.3%)
NUTRITIONAL NEUROSCIENCE	1 (0.9%)	0 (0%)	1 (1.3%)

Osteoporosis International	1 (0.9%)	1 (2.6%)	0 (0%)
PLOS ONE	2 (1.7%)	1 (2.6%)	1 (1.3%)
Preventative Medicine Reports	1 (0.9%)	1 (2.6%)	0 (0%)
Public Health Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
RMD Open	1 (0.9%)	1 (2.6%)	0 (0%)
Scandinavian Journal of Medicine & Science in Sports	6 (5.2%)	6 (16%)	0 (0%)
Scientific Reports	1 (0.9%)	0 (0%)	1 (1.3%)
Social Science & Medicine	1 (0.9%)	0 (0%)	1 (1.3%)
Stroke	1 (0.9%)	1 (2.6%)	0 (0%)
The American Journal of Clinical Nutrition	1 (0.9%)	0 (0%)	1 (1.3%)
The Journal of Gerontology	1 (0.9%)	1 (2.6%)	0 (0%)
The Journal of Nutrition	3 (2.6%)	0 (0%)	3 (3.9%)
The Lancet Regional Health - Western Pacific	1 (0.9%)	0 (0%)	1 (1.3%)
Research area			
Biomechanics	1 (0.9%)	1 (2.6%)	0 (0%)
Clinical nutrition	7 (6.1%)	0 (0%)	7 (9.1%)
Exercise epidemiology	23 (20%)	23 (61%)	0 (0%)
Exercise physiology	1 (0.9%)	1 (2.6%)	0 (0%)
Public health/nutritional epidemiology	70 (61%)	0 (0%)	70 (91%)
Sports medicine	12 (10%)	12 (32%)	0 (0%)
Strength and conditioning	1 (0.9%)	1 (2.6%)	0 (0%)
Study design			
Case-control	5 (4.3%)	0 (0%)	5 (6.5%)
Causal model creation study	3 (2.6%)	1 (2.6%)	2 (2.6%)
Longitudinal cohort	50 (43%)	17 (45%)	33 (43%)
Mendelian randomization	1 (0.9%)	0 (0%)	1 (1.3%)
Randomized controlled trial	2 (1.7%)	1 (2.6%)	1 (1.3%)
Retrospective cross-sectional	52 (45%)	17 (45%)	35 (45%)
Systematic review and/or meta-analysis	2 (1.7%)	2 (5.3%)	0 (0%)

Data source			
1997/1998 São Luís Birth Cohort	3 (2.7%)	1 (2.6%)	2 (2.7%)
Aarhus Birth Cohort (ABC)	1 (0.9%)	1 (2.6%)	0 (0%)
Alberta Pregnancy Outcomes and Nutrition (APrON)	1 (0.9%)	0 (0%)	1 (1.3%)
Assessment of Bisphenol A (BPA) levels and their association with the health status among the Lebanese population	1 (0.9%)	0 (0%)	1 (1.3%)
Australian Health Survey	1 (0.9%)	0 (0%)	1 (1.3%)
Australian Institute of Sport (AIS) customised Athlete Management System	1 (0.9%)	1 (2.6%)	0 (0%)
Australian Longitudinal Study on Women's Health (ALSWH)	4 (3.5%)	3 (7.9%)	1 (1.3%)
Australian MS Longitudinal Study	1 (0.9%)	0 (0%)	1 (1.3%)
Author's own data	15 (13%)	5 (13%)	10 (13%)
BioBank Japan (BBJ)	1 (0.9%)	0 (0%)	1 (1.3%)
BRISA (Brazilian Ribeirão Preto and São Luís Birth Cohort Studies)	2 (1.8%)	1 (2.6%)	1 (1.3%)
Cardiac Wellness Program	2 (1.8%)	2 (5.3%)	0 (0%)
China Birth Cohort Study (CBCS)	1 (0.9%)	0 (0%)	1 (1.3%)
China Multi-Ethnic Cohort (CMEC) study	5 (4.4%)	0 (0%)	5 (6.7%)
Cooperative Health Research in the Region Augsburg (KORA)-Age study	1 (0.9%)	1 (2.6%)	0 (0%)
Danish Diet, Cancer and Health cohort	1 (0.9%)	0 (0%)	1 (1.3%)
Discovery Health and Vitality client data	1 (0.9%)	1 (2.6%)	0 (0%)
Dutch LifeLines cohort	1 (0.9%)	1 (2.6%)	0 (0%)
ELSA-Brasil	1 (0.9%)	0 (0%)	1 (1.3%)
Epidemiological-social study of perinatal health in Ribeirão Preto, São Paulo, Brazil	1 (0.9%)	0 (0%)	1 (1.3%)
Epidemiological surveillance of COVID-19 in the Inconfidentes Region/MG	1 (0.9%)	0 (0%)	1 (1.3%)

EpiFloripa survey	1 (0.9%)	0 (0%)	1 (1.3%)
Étude Longitudinale Française depuis l'Enfance (ELFE)	1 (0.9%)	0 (0%)	1 (1.3%)
Finnish Public Sector (FPS) study	1 (0.9%)	1 (2.6%)	0 (0%)
GarminRUNSAFE Running Health Study and the Coronavirus Resource Centre at John Hopkins University	1 (0.9%)	0 (0%)	1 (1.3%)
Generation R study	1 (0.9%)	0 (0%)	1 (1.3%)
Gretchen Swanson Center for Nutrition (GSCFN)	1 (0.9%)	0 (0%)	1 (1.3%)
Growing Up in Singapore Towards healthy Outcomes (GUSTO)	1 (0.9%)	0 (0%)	1 (1.3%)
IDEFICS/I.Family cohort	1 (0.9%)	0 (0%)	1 (1.3%)
Iranian MS Society (IMSS) and own sample	1 (0.9%)	0 (0%)	1 (1.3%)
Japan Public Health Centerbased (JPHC) diabetes study	1 (0.9%)	0 (0%)	1 (1.3%)
Jinan birth cohort	1 (0.9%)	1 (2.6%)	0 (0%)
Laboratory's historical clinical database collected between 1994 and 2022	1 (0.9%)	1 (2.6%)	0 (0%)
Measuring Eating in everyday Life Study (MEALS)	1 (0.9%)	0 (0%)	1 (1.3%)
Melbourne Collaborative Cohort Study (MCCS)	2 (1.8%)	1 (2.6%)	1 (1.3%)
MESA (Multi-Ethnic Study of Atherosclerosis).	1 (0.9%)	1 (2.6%)	0 (0%)
Mission Locale de Paris	1 (0.9%)	0 (0%)	1 (1.3%)
MoBa-children	1 (0.9%)	0 (0%)	1 (1.3%)
Moli-sani Study	1 (0.9%)	0 (0%)	1 (1.3%)
NFHS and NSS survey	1 (0.9%)	0 (0%)	1 (1.3%)
NHANES	2 (1.8%)	0 (0%)	2 (2.7%)
Niigata Association of Occupational Health in Japan	1 (0.9%)	1 (2.6%)	0 (0%)
No dataset used	1 (0.9%)	0 (0%)	1 (1.3%)
Non-communicable Disease Risk Factor Survey Myanmar 2009	1 (0.9%)	0 (0%)	1 (1.3%)
Norwegian Mother, Father and Child Cohort Study (MoBa)	2 (1.8%)	1 (2.6%)	1 (1.3%)
Norwegian Women and Cancer (NOWAC) cohort	1 (0.9%)	1 (2.6%)	0 (0%)

Nutritionist's Health Study (NutriHS)	2 (1.8%)	0 (0%)	2 (2.7%)
OMEX trial	1 (0.9%)	1 (2.6%)	0 (0%)
ONCODAGE	1 (0.9%)	0 (0%)	1 (1.3%)
Pregnancy Outcomes and Community Health (POUCH) Study	1 (0.9%)	1 (2.6%)	0 (0%)
Pregnancy Study Online (PRESTO) and SnartForældre.dk (SF)	1 (0.9%)	0 (0%)	1 (1.3%)
Pregnancy, Infection, and Nutrition (PIN3) cohort	1 (0.9%)	1 (2.6%)	0 (0%)
PRevention of OVArian Cancer in Quebec (PROVAQ)	1 (0.9%)	0 (0%)	1 (1.3%)
PROMISE-EBF study	1 (0.9%)	0 (0%)	1 (1.3%)
Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial	1 (0.9%)	0 (0%)	1 (1.3%)
RESOLVE project	1 (0.9%)	1 (2.6%)	0 (0%)
Singapore PREconception Study of long-Term maternal and child Outcomes (S-PRESTO)	1 (0.9%)	0 (0%)	1 (1.3%)
Sister Study	1 (0.9%)	0 (0%)	1 (1.3%)
South Asians Living in America (MASALA) study	1 (0.9%)	1 (2.6%)	0 (0%)
Sustainable Undernutrition Reduction in Ethiopia (SURE)	1 (0.9%)	0 (0%)	1 (1.3%)
Swedish CardioPulmonary BioImage Study (SCAPIS).	1 (0.9%)	1 (2.6%)	0 (0%)
Systematic review data	2 (1.8%)	2 (5.3%)	0 (0%)
the 2015 Ethiopia NCD STEPS	1 (0.9%)	0 (0%)	1 (1.3%)
Part of the database of the investigation "Lifestyle, adiposity and vascular function in university students in Castilla-La Mancha, Spain	1 (0.9%)	0 (0%)	1 (1.3%)
The Hispanic Community Health Study/ Study of Latinos (HCHS/SOL)	1 (0.9%)	1 (2.6%)	0 (0%)
The Kuopio Ischaemic Heart Disease (KIHD) Risk Factor study	1 (0.9%)	0 (0%)	1 (1.3%)
The Neomune-NeoNutriNet cohort	1 (0.9%)	0 (0%)	1 (1.3%)
The Nurses' Health Study II (NHS-II)	1 (0.9%)	0 (0%)	1 (1.3%)

The PRevention of OVArian Cancer in Quebec (PROVAQ)	1 (0.9%)	0 (0%)	1 (1.3%)
The Ribeirão Preto birth cohort	1 (0.9%)	0 (0%)	1 (1.3%)
the RPS Consortium Cohorts, Brazil	1 (0.9%)	0 (0%)	1 (1.3%)
The Skin Tumours in Allograft Recipients (STAR) study	1 (0.9%)	0 (0%)	1 (1.3%)
The Stork Groruddalen Study	1 (0.9%)	1 (2.6%)	0 (0%)
The Survey Directed at the Venezuelan Population Residing in the Country (ENPOVE)	1 (0.9%)	0 (0%)	1 (1.3%)
Tianjin Chronic Low-grade Systemic Inflammation and Health (TCLSIH) study	2 (1.8%)	0 (0%)	2 (2.7%)
Tianjin IDD control project and Tianjin thyroid health survey project	1 (0.9%)	0 (0%)	1 (1.3%)
UK Biobank	4 (3.5%)	3 (7.9%)	1 (1.3%)
UK Women's Cohort Study (UKWCS)	1 (0.9%)	0 (0%)	1 (1.3%)
Women's Health Initiative (WHI)	1 (0.9%)	0 (0%)	1 (1.3%)

Presented numbers are counts and proportions for categorical variables, and median and interquartile range for continuous variables.

Supplementary Table S3. Comprehensive list of most common variables across DAG node types

Confounders		Mediators		Outcomes		Exposures		Mediator-outcome confounders	
Variable	**Frequency**	**Variable**	**Frequency**	**Variable**	**Frequency**	**Variable**	**Frequency**	**Variable**	**Frequency**
Age	59 (9.8%)	BMI	23 (13%)	Cancer	6 (6.1%)	Diet	19 (20%)	Alcohol	8 (12%)
Sex	48 (8.0%)	Cardiovascular disease	10 (5.8%)	Mental health	6 (6.1%)	Physical activity	13 (13%)	Physical activity	8 (12%)
Education	34 (5.6%)	Diabetes	10 (5.8%)	Cardiometabolic health	5 (5.1%)	Breastfeeding	4 (4.1%)	Smoking	8 (12%)
Smoking	34 (5.6%)	Energy intake	10 (5.8%)	Diabetes/Glucose Metabolism	5 (5.1%)	Dietary inflammatory index	4 (4.1%)	Health conditions	3 (4.6%)
Income	26 (4.3%)	Dyslipidaemia	9 (5.2%)	Physical Function/Strength	5 (5.1%)	Food security	4 (4.1%)	Stressful events	3 (4.6%)
Physical activity	17 (2.8%)	Adiposity/Obesity	8 (4.6%)	Obesity/Weight Status	4 (4.0%)	Structured Exercise	4 (4.1%)	Age	2 (3.1%)
Socioeconomic status	16 (2.7%)	Physical activity	5 (2.9%)	Pregnancy/ Birth Outcomes	4 (4.0%)	Sugar	4 (4.1%)	Breastfeeding duration	2 (3.1%)
BMI	15 (2.5%)	Age	3 (1.7%)	Child development	3 (3.0%)	Ultra-processed foods	4 (4.1%)	Ethnicity	2 (3.1%)
Alcohol consumption	13 (2.2%)	Cognition/ Mood	3 (1.7%)	Diet	3 (3.0%)	Coffee/Caffeine	2 (2.1%)	Family oral hygiene	2 (3.1%)
Marital status	13 (2.2%)	Mobility and balance	3 (1.7%)	Injurious falls	3 (3.0%)	Dairy	2 (2.1%)	Family sugar consumption	2 (3.1%)
Mental	12 (2.0%)	Muscle	3 (1.7%)	Multiple	3 (3.0%)	Gestational	2 (2.1%)	Khat	2 (3.1%)

health		mass		sclerosis		maternal iron biomarker concentrations		consumption	
Ethnicity	11 (1.8%)	Osteoporosis	3 (1.7%)	Change in pain and function	2 (2.0%)	Muscle mass/Sarcopenia	2 (2.1%)	Last visit to the dentist	2 (3.1%)
Parity	10 (1.7%)	Participation	3 (1.7%)	Early childhood caries	2 (2.0%)	Pregnancy-specific exercise	2 (2.1%)	Meals prepared	2 (3.1%)
Comorbidities/Chronic Diseases	8 (1.3%)	Socioeconomic factors	3 (1.7%)	Mental health	2 (2.0%)			OW/OB	2 (3.1%)
Diet	8 (1.3%)			Mortality	2 (2.0%)				
Maternal age	8 (1.3%)			Periodontal disease status	2 (2.0%)				
Maternal education	8 (1.3%)								
Sleep	6 (1.0%)								
Family disease history	5 (0.8%)								
Menopause	5 (0.8%)								
Profession	5 (0.8%)								
Total energy intake	5 (0.8%)								
Body weight	4 (0.7%)								

Indoor air pollution	4 (0.7%)								
Intake of sweeten beverage	4 (0.7%)								
Maternal schooling	4 (0.7%)								

Reporting was capped at a minimum count of four or three for confounders and mediators, respectively, and two for all other nodes.

Supplementary Table S4. Complete information on statistical methods

Statistical model utilized	**Overall, N = 114**	**Exercise science, N = 38**	**Nutrition, N = 76**
GLM, linear	32 (28%)	13 (34%)	19 (25%)
GLM, logistic	32 (28%)	7 (18%)	25 (33%)
Cox regression	16 (14%)	6 (16%)	10 (13%)
GLM, poisson	8 (7.0%)	2 (5.3%)	6 (7.9%)
Multinomial logistic regression	5 (4.4%)	3 (7.9%)	2 (2.6%)
Linear mixed models	4 (3.5%)	1 (2.6%)	3 (3.9%)
Mixed logistic regression	2 (1.8%)	0 (0%)	2 (2.6%)
Ordinal logistic regression	2 (1.8%)	1 (2.6%)	1 (1.3%)
ANCOVA	1 (0.9%)	0 (0%)	1 (1.3%)
GLM, linear, with mediation analysis via bootstrapping	1 (0.9%)	1 (2.6%)	0 (0%)
GLM, logistic with firth penalization	1 (0.9%)	0 (0%)	1 (1.3%)
GLM, negative binomial	1 (0.9%)	0 (0%)	1 (1.3%)
GLM, polynomial	1 (0.9%)	0 (0%)	1 (1.3%)
Inverse variance weighted model	1 (0.9%)	0 (0%)	1 (1.3%)
Linear mixed model, poisson family	1 (0.9%)	0 (0%)	1 (1.3%)
Mediation analysis from IPD	1 (0.9%)	1 (2.6%)	0 (0%)
Meta-analysis on hazard ratios	1 (0.9%)	1 (2.6%)	0 (0%)
Mixed logistic regression, log binomial family	1 (0.9%)	0 (0%)	1 (1.3%)
Parametric g-formula	1 (0.9%)	1 (2.6%)	0 (0%)
Partial correlation analysis	1 (0.9%)	1 (2.6%)	0 (0%)
Proportional hazards model	1 (0.9%)	0 (0%)	1 (1.3%)
Treatment-outcome relationship	**Overall, N = 112**	**Exercise science, N = 36**	**Nutrition, N = 76**
Categorical regression	75 (67%)	19 (53%)	56 (74%)
Continuous linear	25 (22%)	12 (33%)	13 (17%)

Continuous with RCS	8 (7.1%)	4 (11%)	4 (5.3%)
Continuous polynomial with categorical interaction	1 (0.9%)	0 (0%)	1 (1.3%)
Continuous linear with interaction	2 (1.8%)	1 (2.8%)	1 (1.3%)
Continuous with polynomial	1 (0.9%)	0 (0%)	1 (1.3%)
Adjustment method	**Overall, N = 110**	**Exercise science, N = 35**	**Nutrition, N = 75**
Covariate adjustment in multivariable regression	105 (96%)	34 (97%)	71 (96%)
Inverse probability weighting	3 (2.8%)	0 (0%)	3 (4.1%)
Parametric g-formula	1 (0.9%)	1 (2.9%)	0 (0%)
No statistical analysis performed	1 (0.9%)	0 (0%)	1 (1.3 %)

Presented numbers are counts and proportions. GLM = generalized linear model; ANCOVA = Analysis of covariance; IPD = individual participant data; RCS = restricted cubic spline.