



Developing the WCRF International/University of Bristol Methodology for Identifying and Carrying Out Systematic Reviews of Mechanisms of Exposure–Cancer Associations

Sarah J. Lewis^{1,2}, Mike Gardner^{1,3}, Julian Higgins¹, Jeff M.P. Holly⁴, Tom R. Gaunt^{1,2}, Claire M. Perks⁴, Suzanne D. Turner⁵, Sabina Rinaldi⁶, Steve Thomas⁷, Sean Harrison^{1,2}, Rosie J. Lennon¹, Vanessa Tan^{1,2}, Cath Borwick^{1,8}, Pauline Emmett¹, Mona Jeffreys¹, Kate Northstone⁹, Giota Mitrou¹⁰, Martin Wiseman¹⁰, Rachel Thompson¹⁰, and Richard M. Martin^{1,2}

Abstract

Background: Human, animal, and cell experimental studies; human biomarker studies; and genetic studies complement epidemiologic findings and can offer insights into biological plausibility and pathways between exposure and disease, but methods for synthesizing such studies are lacking. We, therefore, developed a methodology for identifying mechanisms and carrying out systematic reviews of mechanistic studies that underpin exposure–cancer associations.

Methods: A multidisciplinary team with expertise in informatics, statistics, epidemiology, systematic reviews, cancer biology, and nutrition was assembled. Five 1-day workshops were held to brainstorm ideas; in the intervening periods we carried out searches and applied our methods to a case study to test our ideas.

Results: We have developed a two-stage framework, the first stage of which is designed to identify mechanisms underpinning a

specific exposure–disease relationship; the second stage is a targeted systematic review of studies on a specific mechanism. As part of the methodology, we also developed an online tool for text mining for mechanism prioritization (TeMMPo) and a new graph for displaying related but heterogeneous data from epidemiologic studies (the Albatross plot).

Conclusions: We have developed novel tools for identifying mechanisms and carrying out systematic reviews of mechanistic studies of exposure–disease relationships. In doing so, we have outlined how we have overcome the challenges that we faced and provided researchers with practical guides for conducting mechanistic systematic reviews.

Impact: The aforementioned methodology and tools will allow potential mechanisms to be identified and the strength of the evidence underlying a particular mechanism to be assessed. *Cancer Epidemiol Biomarkers Prev*; 26(11); 1–9. ©2017 AACR.

Introduction

Systematic reviews offer robust methodology for identifying, appraising, and synthesizing studies that have addressed a common research question (1, 2). Such reviews are valuable in the synthesis of published literature relating to health care interventions and to etiologic questions. However, reviews of observational epidemiologic findings by themselves are insufficient to establish causation. Other forms of evidence are required to complement such data to infer the likely causality of any observed association, in particular biological plausibility (3). There is an abundance of evidence relating to the biology underpinning the causation of disease, from studies such as human, animal, and cell experimental studies; human biomarker studies; and genetic association studies, although methods have not been developed to synthesize this in a systematic way. Consequently, although epidemiologic studies addressing chronic disease can be synthesized using a systematic process, mechanistic studies have previously been addressed using a results narrative.

The World Cancer Research Fund (WCRF) and American Institute for Cancer Research have published a landmark report addressing the prevention of cancer through diet, nutrition, and physical activity (4). As part of the Continuous Update of the 2007

¹Bristol Medical School, Population Health Sciences, University of Bristol, Bristol, United Kingdom. ²MRC Integrative Epidemiology Unit (IEU) at the University of Bristol, Bristol, United Kingdom. ³Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom. ⁴IGFs & Metabolic Endocrinology Group, Learning and Research Building, Bristol Medical School, Translational Health Sciences, Southmead Hospital, Bristol, United Kingdom. ⁵Department of Pathology, University of Cambridge, Cambridge, United Kingdom. ⁶International Agency for Research on Cancer, Lyon, France. ⁷Bristol Dental School, University of Bristol, Bristol, United Kingdom. ⁸Cardiff University, Cardiff, United Kingdom. ⁹CLAHRC West, University of Bristol, Bristol, United Kingdom. ¹⁰WCRF International, London, United Kingdom, Bristol, United Kingdom.

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org/>).

Corresponding Author: Sarah J. Lewis, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, United Kingdom. Phone: 796-641-0953; Fax: 117-928-7325; E-mail: s.j.lewis@bristol.ac.uk

doi: 10.1158/1055-9965.EPI-17-0232

©2017 American Association for Cancer Research.

Report (5), WCRF UK commissioned the University of Bristol to develop a framework for reviewing mechanistic studies of exposures and cancer to test the likely causality of the observed associations. The aims were to (i) identify mechanistic studies that provide evidence of the biological plausibility of the causality of links between a diet, nutrition or physical activity exposure, and cancer; and ii) systematically review and assess the strength of the evidence for any one particular mechanism.

Challenges in conducting systematic reviews of the mechanisms mediating observed associations between potentially modifiable exposures and cancer

- (i) How to identify the relevant mechanisms for a particular exposure–outcome association
- (ii) How to cope with the enormous wealth of data generated in searching for mechanisms
- (iii) How to assess the quality of animal and cell studies
- (iv) How to determine the relevance of animal studies to human disease
- (v) How to assess the extent of publication bias
- (vi) How best to integrate all the evidence

We outline how we addressed the challenges inherent in developing an overall methodology outlined above. A schematic diagram of the steps is given in Fig. 1 with full details of the methodology presented in the Supplementary Material.

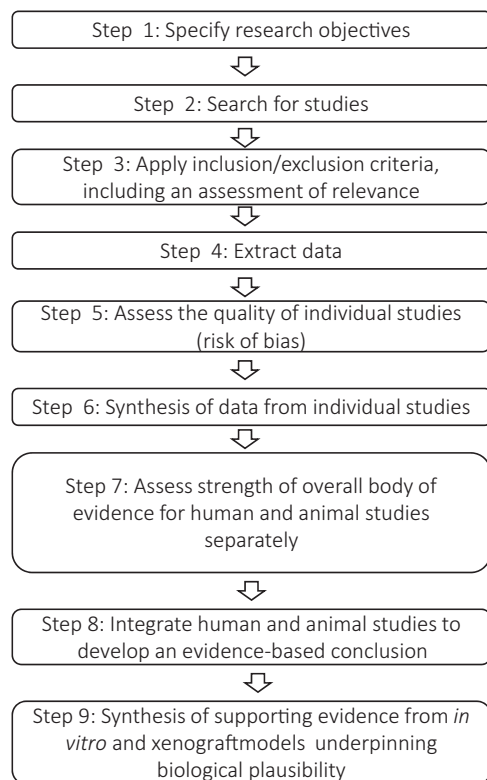


Figure 1.

Steps for stage 2. Figure 1 shows an outline of the steps we recommend going through in stage 2 of our methodology to review the evidence for a specific mechanism.

Materials and Methods

We approached colleagues and collaborators from the University of Bristol (Bristol, United Kingdom), University of Cambridge (Cambridge, United Kingdom), and the International Agency for Research on Cancer (Lyon, France) to assemble a multidisciplinary team with expertise in bioinformatics (T.R. Gaunt), statistics (J. Higgins, S. Harrison, K. Northstone, and R.M. Martin), cancer biology (J.M.P. Holly, C.M. Perks, and S. Thomas), animal studies (J.M.P. Holly, M. Gardner, and S. Thomas), molecular biology (T.R. Gaunt, J.M.P. Holly, C.M. Perks, V. Tan, and S. Thomas), epidemiology (S.J. Lewis, P. Emmett, M. Jeffreys, K. Northstone, and R.M. Martin), genetic epidemiology (S.J. Lewis and T.R. Gaunt), nutrition (S. Rinaldi, P. Emmett, and K. Northstone), and systematic reviews (S.J. Lewis, M. Gardner, J. Higgins, and R.M. Martin). Our objective to develop a rigorous systematic review methodology integrating animal, cell, and human studies was met through a combination of discussion workshops and advice from a panel of experts. Decisions were reached by discussion and consensus opinion and then tested in practice. Results were fed back to the team, and changes were made to the methodology if needed.

We tested the framework by implementation in a case study examining the IGF pathway to determine whether this could explain observed associations between consumption of milk and incidence of prostate cancer (reported in full separately). To do this, we systematically reviewed evidence on milk intake and the IGF pathway, and between the IGF pathway and prostate cancer (6). In this review, we pooled together evidence from randomized controlled trials and other experimental studies in humans, observational, human biomarker, genetic, and animal studies. The feasibility and reproducibility of our methodology has been independently tested by two teams of systematic reviewers who initially searched for mechanisms between higher body fatness and postmenopausal breast cancer, and systematically reviewed the insulin-like growth factor 1 receptor as a potential mechanism for this association (7). The findings by Ertaylan and colleagues are published as an article in the same issue of this journal (7).

Results

Identifying the relevant mechanisms for a particular exposure–outcome association

We have developed a two-stage strategy: In stage 1 all potential mechanisms underlying a particular exposure–outcome association are identified, taking a largely "hypothesis-free" approach; in stage 2, the evidence underlying one or more specific mechanisms is systematically reviewed. Fundamental to our approach are "intermediate phenotypes" (IP) between the exposure and disease (e.g., measures of DNA damage) as mechanistic studies frequently have an IP rather than cancer as an outcome, or will investigate the IP as the exposure in relation to an outcome. Stage 1 assembles the evidence around IPs, to determine which have evidence linking them either to the exposure or to the outcome, and to quantify this evidence. For the study of milk and prostate cancer, a list of potential IPs was generated (Table 1). In doing this, we considered the biological processes that may lead to prostate cancer, referring to important reviews in the area of cancer, such as those on the hallmarks of cancer (8), which have been proposed as a framework for considering disordered biology in malignancies. In addition, reviews specific to the cancer site (in our case prostate cancer) were consulted to identify potential mechanisms. General

Table 1. Intermediate phenotypes used in a review of milk and prostate cancer

MeSH Terms (in bold) and more specific terms (nonbold)	
Nerve growth factors	Receptors, steroid
Brain-derived neurotrophic factor	Bone marrow
Ciliary neurotrophic factor	Enterochromaffin cells
Glia maturation factor	Immunologic synapses
Glial cell line-derived neurotrophic factors	Leukocytes
Nerve growth factor	Lymphatic system
Neuregulins	Mast cells
Neurotrophin 3	Phagocytes
Pituitary adenylate cyclase-activating polypeptide	Mononuclear phagocyte system
Membrane transport proteins	Angiogenesis-modulating agents
ATP-binding cassette transporters	Angiogenesis-inducing agents
Amino acid transport systems	Angiogenesis inhibitors
Fatty acid transport proteins	Signal transduction
Ion channels	Ion channel gating
Ion pumps	Light signal transduction
Monosaccharide transport proteins	MAP kinase signaling system
Neurotransmitter transport proteins	Mechanotransduction, cellular
Nucleobase, nucleoside, nucleotide, and nucleic acid transport proteins	Second messenger systems
Nucleocytoplasmic transport proteins	Synaptic transmission
Racemases and epimerases	Energy metabolism
Amino acid isomerases- alanine racemase	Basal metabolism
Carbohydrate epimerases- UDPglucose 4-epimerase	Citric acid cycle
Glutathione transferase	Glycolysis
Glutathione S-transferase pi	Oxidation-reduction
Androgens	Oxidative phosphorylation
Dihydrotestosterone	Pentose phosphate pathway
Nandrolone	Photophosphorylation
Oxandrolone	Proton-motive force
Oxymetholone	Substrate cycling
Stanozolol	Cell differentiation
Testosterone	Adipogenesis
Androgen antagonists	Asymmetric cell division
Chlormadinone acetate	Embryonic induction
Cyproterone	Gametogenesis
Cyproterone acetate	Hematopoiesis
Flutamide	Neurogenesis
Transactivators	Cell death
Gene products, tat	Apoptosis
Herpes simplex virus protein Vmw65	Autophagy
Very broad/general MeSH terms not subdivided for more specific terms	Necrosis
Receptors, androgen	
Receptors, estrogen	
Receptors, glucocorticoid	
Receptors, mineralocorticoid	
Receptors, progesterone	
Molecular mechanisms	
Physiology	
Cell physiologic processes	
Genomic instability	MeSH terms without more specific terms
Chromosomal instability- chromosome fragility	Selenium
Microsatellite instability	miRNAs
DNA damage	DNA methylation
DNA adducts	C-Reactive Protein
DNA breaks—chromosome breakage	Telomerase
DNA degradation, necrotic	
DNA fragmentation	Hormones and growth factors (title—not MeSH term)
DNA repair	Testosterone
DNA end-joining repair	Estrogens
DNA mismatch repair	
Recombinational dna repair	Somatomedins
SOS response	Insulin-like growth factor i
Gene expression	Insulin-like growth factor ii
Protein biosynthesis	Insulin-like growth factor binding proteins
Transcription, genetic—reverse transcription; transcriptome	Insulin-like growth factor binding protein 1
Mutation	Insulin-like growth factor binding protein 2
Allelic imbalance	Insulin-like growth factor binding protein 3
Base pair mismatch	Insulin-like growth factor binding protein 4
Chromosome aberrations	Insulin-like growth factor binding protein 5
Codon, nonsense	Insulin-like growth factor binding protein 6
DNA repeat expansion	
	Vitamins and minerals (title—not MeSH term)
	Calcium, dietary
	Vitamin D

(Continued on the following page)

Table 1. Intermediate phenotypes used in a review of milk and prostate cancer (Cont'd)

Mutagenesis	
Frameshift mutation	Amino acid substitution sequence inversion
Gene amplification	Chromosome duplication
Gene duplication	Nondisjunction, genetic
Germline mutation	Somatic hypermutation, immunoglobulin
INDEL mutation	Translocation, genetic
Mutagenesis, insertional	Genomic instability
Mutation rate	Chromosomal instability—chromosome fragility
Mutation, missense	Suppression, genetic
Point mutation	Microsatellite instability
Sequence deletion	Terms entered as title not MeSH terms
Cytokines	Inflammation
Chemokines	Immunity
Growth differentiation factor 15	Programmed cell death
Hematopoietic cell growth factors	Physiology programmed cell death
Hepatocyte growth factor	Prostatitis physiology
IFNs	Physiology prostatitis
IL1 receptor antagonist protein	Prostatitis physiology
Interleukins	Prostatitis
Leukemia inhibitory factor	
Lymphokines	
Monokines	
Oncostatin M	
Osteopontin	
TGF β	
TNFs	
Cell proliferation	
Cell division—asymmetric cell division; telomere homeostasis	
Immune system	
Antibody-producing cells	
Antigen-presenting cells	

MeSH terms relating to potential IPs were used in the search whenever possible, rather than more specific terms, as this allowed a broader search to be carried out. Reviewers can generate their own list of IPs by listing terms relating to general cancer processes (such as the hallmarks of cancer), searching for reviews on the biology of their cancer site of interest and seeking expert opinion. We would advocate being as inclusive as possible at this point.

Coping with the enormous wealth of data that is generated in searching for mechanisms

The sheer number of articles generated in stage 1 (>39,000 in our case study of milk and prostate cancer) meant that we needed an efficient strategy for processing these data and prioritizing mechanisms for full systematic review in stage 2. Therefore, we have devised an automated process ["Text Mining for Mechanism Prioritisation" (TeMMPo)] that allows quantification and visualization of the amount of evidence underlying each step in the mechanistic pathway ($E \rightarrow IP$, $IP \rightarrow C$, $E \rightarrow C$, where E is exposure, IP is intermediate phenotype, and C is cancer). This tool can be accessed at <https://www.temmpo.org.uk/>. The program allows users to upload the results of their MEDLINE or PubMed searches, which are then displayed according to the intermediate phenotypes in a Sankey plot. This illustrates the quantity of evidence linking specific IPs with exposures ($E \rightarrow IP$) and the quantity of evidence linking the same IPs with disease ($IP \rightarrow C$); the relative number of publications underlying each link is depicted by the thickness of the lines linking the terms. A weighted score is generated as follows: the number of publications for E-IP or IP-C (whichever is the least) divided by the number of publications for E-IP or IP-C

(whichever is the greater) multiplied by the total number of publications for each intermediate phenotype. According to this score, IPs are then ranked. These data then inform the selection of specific intermediates to be investigated in stage 2. Figure 2 shows a Sankey plot generated by TeMMPo indicating the quantity of studies linking milk with an IP and the quantity of studies linking the same IP with a prostate cancer outcome.

The limitations of this approach are: it assumes that the cooccurrence of a biological mechanism with exposure or outcome in the literature represents an association rather than simply a cooccurrence of the two terms in the same article; it assumes the mechanisms are represented by a single mediating factor; recently identified pathways will be underrepresented in this approach as they are likely to have fewer studies; and it does not address issues of study type, quality, direction, and magnitude of results.

Systematically reviewing the evidence for a particular mechanism including assessing study quality

Having identified potential mechanisms underlying a particular exposure–outcome association, stage 2 systematically reviews the evidence underlying one or more specific mechanisms. For our study of milk–prostate cancer, we chose to systematically review the IGF pathway, as our stage 1 searches indicated that on combining all related IP terms, there were more studies linking IGF intermediates (i.e., a combination of IGF-I, IGF-II, IGF-IR, IGFBP3, IGFBP1) with both milk and prostate cancer than for other potential mechanisms.

Stage 2 largely follows standard systematic review methodology (see Supplementary Material): specification of research

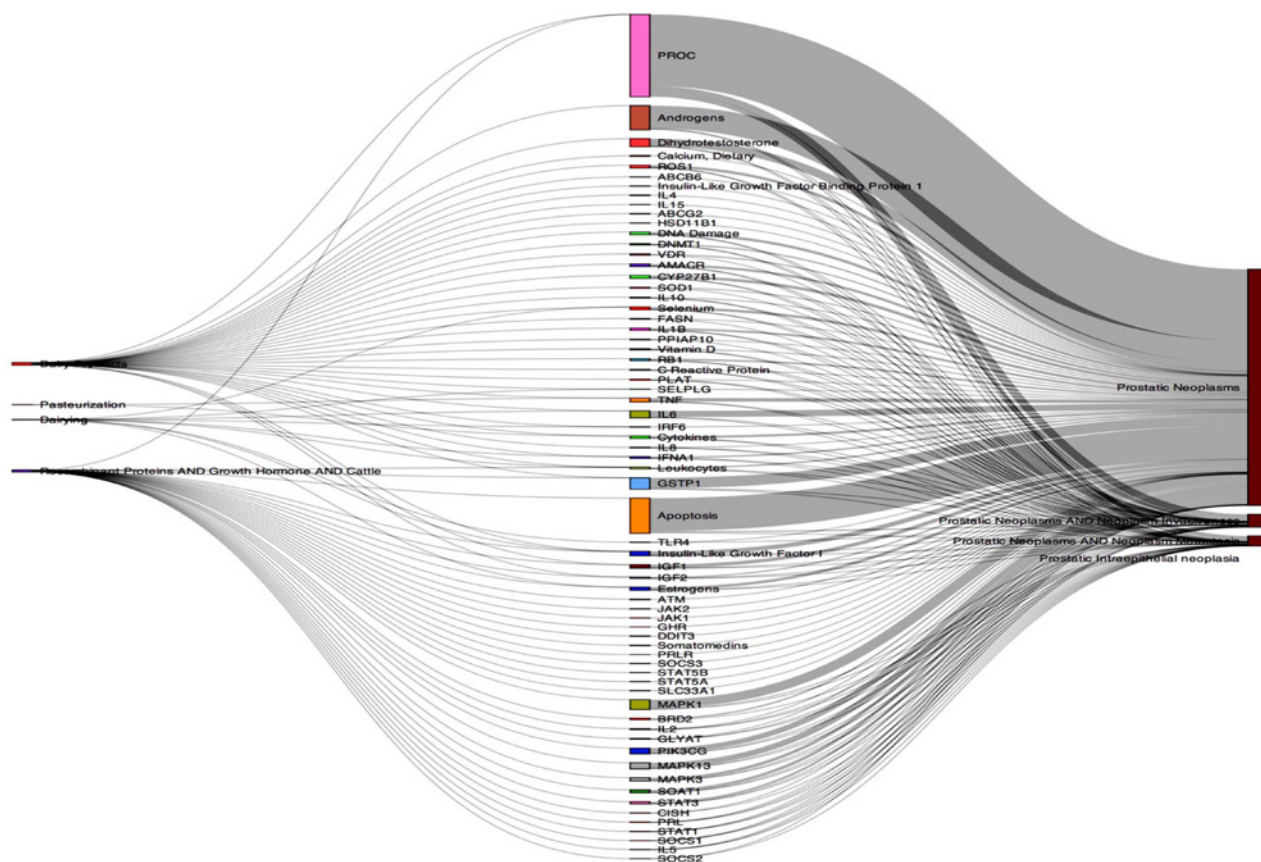


Figure 2.

A Sankey plot of milk-IGF-prostate cancer. Figure 2 shows a Sankey plot that indicates visually the quantity of evidence linking exposure to different intermediate phenotypes and the quantity of evidence linking the same intermediate phenotypes to outcome. This particular Sankey plot shows the quantity of evidence for milk and IGF on the left-hand side and the quantity of evidence for IGF–prostate cancer on the right-hand side of the plot.

objectives; conduct searches (see Supplementary Table S1 as a guide for developing search terms); apply inclusion/exclusion criteria; extract data; assess study quality and synthesize data across studies. Existing tools for assessing study quality have not been validated or established for mechanistic (9–11) nor animal studies (12). We recommend the Cochrane risk of bias tools for human studies (9) and SYRCLE (Systematic Review Centre for Laboratory animal Experimentation; ref. 13), which adapts the Cochrane tool (9), for aspects of bias that are specific to animal studies. SYRCLE addresses the following domains:

- Bias due to confounding (sequence generation, baseline characteristics, allocation concealment)
- Bias due to departures from intended intervention (e.g., due to lack of random housing of animals or lack of blinding)
- Bias due to missing data
- Bias in measurement of outcomes
- Bias in selection of reported results

As far as we are aware, there are currently no tools for assessing the quality of cell line studies, so we developed the criteria listed as follows through consensus of the framework development group, which included cell biologists. Supplementary Table S2 recommends variables to extract by study type at data extraction stage to complete the risk of bias assessments.

Criteria used for assessing the quality of cell studies

- Have the cells been obtained from a validated repository that guarantees cell verification or have the cells been appropriately independently verified?
- Have sufficient biological and technical repeats of the experiments been conducted and were appropriate controls included?
- Were different cell lines from the same cancer type used in the study? An effect observed in more than just one cell line implies the effect is important and relevant to this cancer type.
- Are culture conditions comparable between different studies?
- Selective reporting: are only selected results from several cell line experiments reported?
- Were cell lines from different cancer types compared? This implies an important effect that is relevant more generally to cancer cells.

We recommend that questions 1 to 3 above are used to determine inclusion of cell studies into the review. In our study of milk-IGF-prostate cancer, only a small proportion of relevant cell studies met these basic quality criteria (Fig. 3). However, it is a recent requirement to provide authentication of cell lines

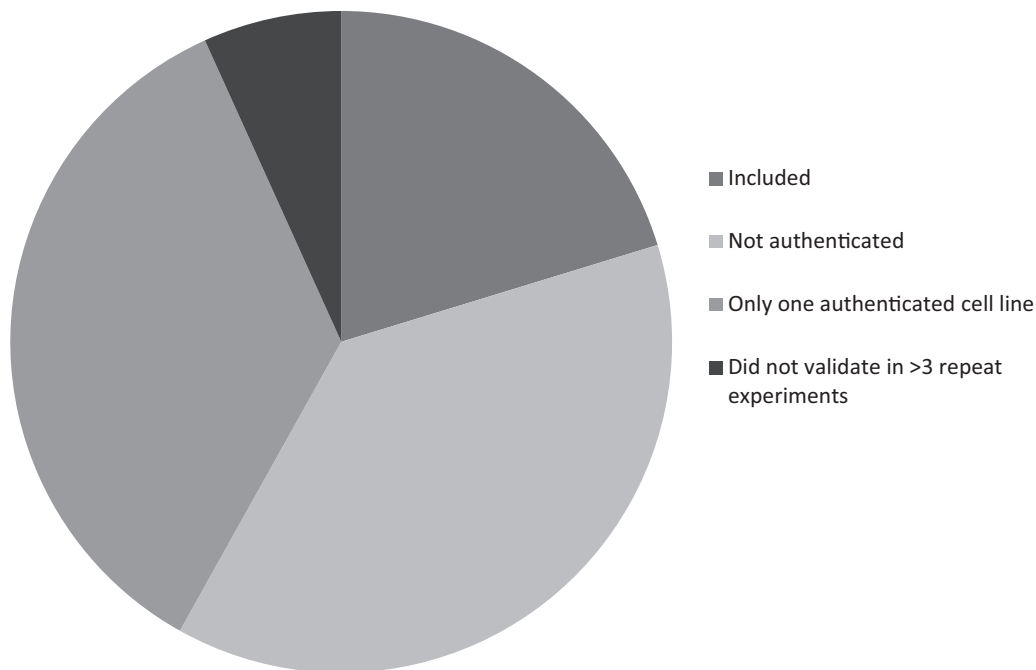


Figure 3.

Pie chart showing proportion of cell studies included after applying quality control criteria and reasons for exclusion in our study of milk-IGF-prostate cancer. Figure 3 shows that our search identified 74 articles of cell studies relevant to milk-IGF-prostate cancer the research question; of these, 59 were excluded because they did not use authenticated cell lines ($n = 28$), carried out experiments in only one authenticated cell line ($n = 26$), or did not validate results in more than three repeat experiments ($n = 5$).

and other quality control criteria for publication. Thus, in applying these criteria, we are selecting more recent studies and may be excluding high-quality historical studies, which were not required to provide information on the above to publish. Questions 4 to 6 can be used to assess the reproducibility of the findings from cell studies.

Synthesis of individual studies and "Albatross plots" for graphical representation of evidence synthesis, when meta-analysis is not appropriate

The next step is the synthesis of data from individual studies. Formal meta-analysis of comparable studies is recommended where possible and appropriate (14). However, it is likely that mechanistic studies will be too heterogeneous (in terms of exposure and outcome definitions; different follow-up periods; different study types) to combine, and therefore, some studies will only be amenable to a narrative summary of the results. We therefore developed a new method to graphically represent heterogeneous data, which we have termed "Albatross plots" (15). These plots allow for the strength and direction of association to be displayed continuously, plotting P values against the number of participants in the studies (which will give an indication of the relative power of the study; Fig. 4). Clustering of data points toward one side of the graph represents an association between exposure and outcome in that direction. In Fig. 4, the majority of studies are on the right side of the graph, indicating a positive association of exposure (milk and dairy products) with outcome (IGF-I). Small studies will only have low P values if the effect size is large, whereas large studies may have low P values even when the effect size is small.

Contour lines that indicate a specific β -coefficient can be added to the plot to indicate (to some extent) the magnitude of association. Simple contours can be computed on the basis of P values and the number of participants, although it should be noted that such contours are not sufficient or appropriate to provide a precise effect estimate (as a forest plot would). Contours can be added if the majority of data have been analyzed in the same way (linear or logistic regression, or standardized mean differences), and the contour will be of the same type of effect estimate (e.g., a standardized β -coefficient for linear regression). If data points fall along a contour (which is shaped like a bird's wing, hence "Albatross plots"), then there is likely to be an association of the magnitude represented by the contour; however, this needs to be interpreted with a narrative and consideration of the individual studies in the synthesis.

We did not find any animal or cell studies that addressed the association between milk and IGF intermediates, but the 8 animal studies on IGF-prostate cancer outcomes were too varied (different experiments, on alternative aspects of the IGF pathway, in diverse animal models, with varied outcomes), to combine in a plot. Characteristics and results of these studies were tabulated (see ref. 6). A schematic diagram of the likely biological pathway generated from animal and cell line studies is another way of presenting the data.

Assessment of the strength of evidence and classification of studies according to relevance to humans

Once the synthesis of evidence has been completed, the framework requires an assessment of the strength of the body of evidence. We recommend doing this separately for human and

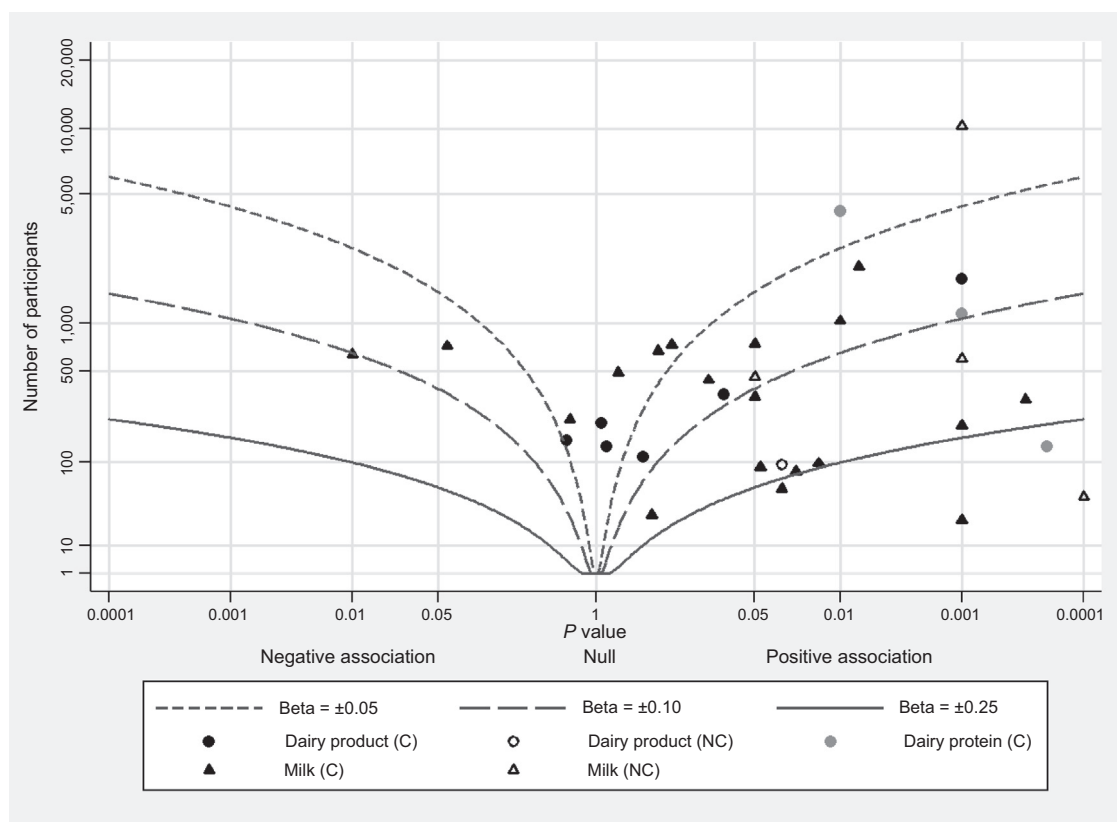


Figure 4.

Albatross plot of milk, dairy products and dairy proteins (exposures), and IGF-I (outcome). Figure 4 shows that the majority of studies are on the right side of the graph, indicating a positive association of exposure with outcome. Note also that the majority of studies showing an association do so around a standardized beta coefficient (Beta) of 0.1, which is a 0.1 SD increase in outcome for a 1 SD increase in exposure.

animal studies, according to the GRADE framework (16), which has been adopted by the Cochrane Collaboration.

Although our remit was to design a framework that could be used to incorporate relevant evidence from any type of study, some studies were so far removed from humans that they could not inform a judgment that a particular process is operating in the human disease pathway. However, such studies could be used to assess general biological plausibility. For cancer, we chose to distinguish between two types of animal models by applying the question "Has the cancer arisen *de novo* in the animal model rather than being transplanted into the animal?" This is because transplantable models represent cancers that are already highly evolved as they have adapted growth *in vitro* (in the case of cell line xenografts) or *in vivo* growth in patient-derived xenograft models (human tumor cells taken from host patient and transplanted into immunodeficient mice) and are typically of a more aggressive biological phenotype; as such, they do not closely mimic most human cancers and are unlikely to give useful information about the usual process of cancer development or progression.

We recommend that only studies that closely mimic human cancers should be used to determine the strength of the evidence underlying a particular mechanistic pathway in human cancer. Other animal studies could be assessed alongside cell line studies to determine whether they provide evidence for the general biological plausibility of the proposed mechanism.

In addition to this two-tiered distinction when applying the GRADE framework, studies are assessed according to the following criteria: indirectness (this relates to how well the study addresses the specific research question), inconsistency, imprecision, and publication bias.

As we are not aware of the GRADE framework being previously applied to animal studies, the question of indirectness in particular required some consideration. We therefore developed some questions to assess this specifically for animal studies.

Assessing the indirectness of animal studies when applying the GRADE framework

- Is the exposure applied via a route that is comparable with that in humans, and a mode that addresses the research question? (e.g., if the interest is in a food exposure, then this should be ingested by the animal model; for other exposures, it may be appropriate to introduce this via an alternative route).
- Is the level and frequency of exposure comparable with that which humans may experience after accounting for species differences in pharmacokinetics and pharmacodynamics, or is the dose justified within the study? (much greater doses than would be possible or reasonable in humans are unlikely to reflect human exposures)

- Is the cancer induced (i.e., by a virus, radiation, chemical agent, or genetic manipulation; whether or not these studies can be included will depend on the research question, but the agent used should be relevant to the human cancer)?
- Is the time at which the outcome is assessed justified? Whether the timing of outcome assessment is relevant will depend on the outcome, e.g., if the outcome is a gene mutation then that outcome could justifiably be assessed very quickly following exposure, but if the outcome is cancer this may require much longer follow-up to produce relevant data.
- Does the study explore mechanisms or pathways of cancer development?
- Is the outcome of assessment cancer incidence or progression rather than surrogate measures of tumor activity such as tumor size or number of tumors?
- Do the outcome measures mimic those found in humans? More specifically, does the tumor mimic the human disease in terms of the organ or tissue affected, and at the histopathologic (tissue patterns, or cell surface, or intracellular protein expression levels) or genetic level (are equivalent hallmark genetic lesions observed as well as gene expression profiles)? Does the progression of the disease mimic the human cancer (e.g., metastasis to the same sites, vascular and stromal invasion, response to treatment)?

If the answer to one or more of these questions is no, then the individual study should be considered to offer indirect evidence; if the majority of studies in the body of evidence are considered to offer only indirect evidence, then the overall GRADE assessment across these studies should be downgraded. For example, we downgraded animal studies of IGF and prostate cancer because knock-out mice do not represent variation within the normal range, and in some studies, the outcome measured was tumor weight or volume rather than incidence.

Investigating whether publication bias is likely to have occurred

There is empirical evidence that studies with null results (no association) are less likely to be in the published literature. Null studies may also be affected by "time lag bias" or longer time to publication. Funnel plots and the Begg (17) and Egger (18) tests can be used to examine for association between effect sizes and study sizes (essentially sample size), and such an association ("small study effect") may reflect publication bias. However, these approaches may not be possible due to an insufficient number of similar studies with the same exposures and outcomes measured. Ioannidis and Trikalinos (19) have developed a method to test for excess statistical significance across studies on different research questions within the same domain. Domains may be defined according to a common general theme, intervention type, subject type, methodology, research environments, and language of publication or combinations of these factors. The test is a comparison of the number of observed studies with statistically significant results compared against the number of expected statistically significant results among all meta-analyses considered in the domain. This test can be applied to assess publication bias across domains.

An alternative approach is to qualitatively assess publication bias by obtaining data on unpublished studies (e.g., by searching the gray literature and/or contacting researchers working in the field) to determine whether relevant unpublished experiments or observational studies have been carried out. It is difficult to be

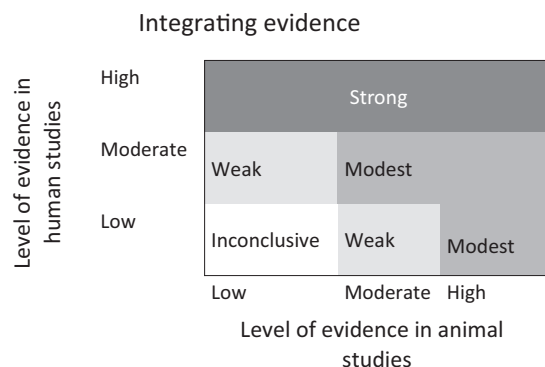


Figure 5.

A guide to integrating the evidence from human and animal studies to reach an overall conclusion on the strength of evidence for a particular mechanism underlying an exposure and cancer association. Figure 5 shows how overall conclusion on the strength of evidence for exposure-intermediate and intermediate-outcome may be reached on the basis of evidence from animal and human studies. This was adapted from the National Toxicology Program (20).

systematic about such investigations, but attempts should be fully reported to ensure transparency of the process. Reviewers can then compare the results of any unpublished or gray literature studies with those that have been published to determine whether there are important differences in the results. This process may indicate non-, delayed, or restricted (e.g., in difficult-to-retrieve journals) publication of null data, suggesting distortion of the mainstream literature by publication bias.

Assessing the strength of evidence across evidence streams and synthesis of cell line and other animal studies

In the WCRF International/University of Bristol framework (Supplementary Material), we have set out a model for assessing the totality of evidence by determining the strength of the overall evidence from human and animal studies, which reflect the human disease process (see Fig. 5). In addition, we advocate using other studies to illustrate biological plausibility and illustrate the potential intricacies of the biological pathway.

Discussion

We have developed a methodology that can be used to identify potential mechanisms underlying observed associations between an exposure and an outcome and to systematically review a mechanistic pathway of interest. We have overcome several hurdles, including developing an automated online tool (<https://www.temmpo.org.uk/>) to deal with the vast amounts of studies identified in stage 1; recommending tools for assessing the quality and relevance of animal and cell studies to human disease; and developing a new method for synthesizing data from a variety of study types, the Albatross plot. However, implementing the methodology does have some limitations, the main one being that it is very time consuming, which may constrain its use. In addition, we have seen from our case study that many animal and cell studies do not report basic information that we recommend using to assess their quality; this is particularly true for older research findings. This means that many studies that are pertinent to the research question may not be included in the overall

analysis. Furthermore, there is a question mark over the relevance of animal experiments to the human situation, although we have made suggestions for assessing how relevant they may be and for weighting these studies accordingly in the overall analysis.

We believe that the methodology we have developed can be applied to the integration of mechanistic studies into systematic reviews of exposures and disease to aid the inference of causality, and in addition may highlight gaps in our knowledge where further studies are needed.

Disclosure of Potential Conflicts of Interest

T.R. Gaunt reports receiving commercial research grants from Biogen, GlaxoSmithKline, and Sanofi. S.D. Turner reports receiving a commercial research grant from GlaxoSmithKline. No potential conflicts of interest were disclosed by the other authors.

Authors' Contributions

Conception and design: S.J. Lewis, J.M.P. Holly, S.D. Turner, M. Jeffreys, R.M. Martin

Development of methodology: S.J. Lewis, M. Gardner, J. Higgins, J.M.P. Holly, T.R. Gaunt, C.M. Perks, S.D. Turner, S. Thomas, S. Harrison, R.J. Lennon, C. Borwick, P. Emmett, M. Jeffreys, G. Mitrou, M. Wiseman, R.M. Martin

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): R.M. Martin

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.J. Lewis, M. Gardner, J. Higgins, J.M.P. Holly, C.M. Perks, S. Harrison, M. Jeffreys, R.M. Martin

Writing, review, and/or revision of the manuscript: S.J. Lewis, M. Gardner, J. Higgins, J.M.P. Holly, T.R. Gaunt, C.M. Perks, S.D. Turner, S. Rinaldi, S. Thomas, S. Harrison, R.J. Lennon, C. Borwick, P. Emmett, K. Northstone, G. Mitrou, M. Wiseman, R. Thompson, R.M. Martin

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): M. Gardner, T.R. Gaunt, V. Tan, C. Borwick

Study supervision: S.J. Lewis, M. Wiseman, R.M. Martin

Acknowledgments

We would thank the Mechanisms Protocol Development Group (Drs. Andrew Dannenberg, Johanna Lampe, Henry Thompson, Steven Clinton, Stephen Hursting, Nikki Ford) and Dr. Susan Higginbotham (member of Secretariat) who initiated this work and whose protocol we have referred to in developing these guidelines. We would especially like to acknowledge the input of Drs. Stephen Hursting and Steven Clinton in relation to assessing the relevance of animal studies to human disease.

Grant Support

All authors received a grant from the World Cancer Research Fund (grant number: RFA 2012/620). S. Harrison is a Wellcome Trust Funded PhD student, 102432/Z/13/Z. R.M. Martin, S.J. Lewis, J.M.P. Holly, T. Gaunt, C.M. Perks are supported by a Cancer Research UK (C18281/A19169) Programme Grant (the Integrative Cancer Epidemiology Programme).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 15, 2017; revised June 28, 2017; accepted September 19, 2017; published OnlineFirst October 4, 2017.

References

- Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Med* 2009;6:e100097.
- Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000;283:2008–12.
- Hill AB. The environment and disease: association or causation?. *Proc R Soc Med* 1965;58:295–300.
- World Cancer Research Fund/American Institute for Cancer Research. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Washington DC: AICR, 2007.
- World Cancer Research Fund International. Continuous Update Project. Available from: www.wcrf.org/cup.
- Harrison S, Lennon R, Holly J, Higgins JPT, Perks C, Gardner M, et al. Does milk intake promote prostate cancer initiation or progression via effects on insulin-like growth factors (IGFs)? A systematic review and meta-analysis. *Cancer Causes Control* 2017;28:497–528.
- Ertaylan G, Le Cornet C, van Roekel EH, Jung AY, Bours MJL, Damms Machado A, et al. A comparative study on the WCRF International/University of Bristol methodology for systematic reviews of mechanisms underpinning exposure-cancer associations. *Cancer Epidemiol Biomarkers Prev* 2017;26:1583–94.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144:646–74.
- Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;355:i4919.
- Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess the risk of bias in systematic reviews. *J Clin Epidemiol* 2016;69:225–34.
- Krauth D, Woodruff TJ, Bero L. Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. *Environ Health Perspect* 2013;121:985–92.
- Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLEs risk of bias tool for animal studies. *BMC Med Res Methodol* 2014;14:43.
- DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88.
- Harrison S, Jones HE, Martin RM, Lewis S, Higgins JPT. The Albatross plot: a novel graphical tool for presenting results of diversely reported studies in a systematic review. *Res Synthesis Methods* 2017. Available from: <https://doi.org/10.1002/jrsm.1239>.
- Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;50:1088–101.
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ* 1997;315:629–34.
- Ioannidis JPA, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials* 2007;4:245–53.
- National Toxicology Program, U.S. Department of Health and Human Services. Draft OHAT approach for systematic review and evidence integration for literature-based health assessments- February 2013. Washington, DC: U.S. Department of Health and Human Services; 2013.

Cancer Epidemiology, Biomarkers & Prevention

AACR American Association
for Cancer Research

Developing the WCRF International/University of Bristol Methodology for Identifying and Carrying Out Systematic Reviews of Mechanisms of Exposure–Cancer Associations

Sarah J. Lewis, Mike Gardner, Julian Higgins, et al.

Cancer Epidemiol Biomarkers Prev Published OnlineFirst October 4, 2017.

Updated version	Access the most recent version of this article at: doi: 10.1158/1055-9965.EPI-17-0232
Supplementary Material	Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2017/10/04/1055-9965.EPI-17-0232.DC1

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/early/2017/10/20/1055-9965.EPI-17-0232 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.