|  |
| --- |
| WCRF International/University of Bristol |
| A systematic review framework for integrating evidence from human, animal and other mechanistic studies |
|  |
|  |
|  |
| **3/15/2017** |

|  |
| --- |
|  |



### Stage 1- Identifying potential mechanisms by which an exposure causes an outcome

Stage 1 has three steps:

1. Defining the research question;
2. Conducting a literature search; and
3. Analysing the findings from the search.

Our approach in this stage is highly inclusive, and the search may identify a very large number of articles. We therefore adopt automated approaches to summarising the literature, making use of text mining methods (Stage 1, Step 3).

*This stage may be bypassed if the objective is to review the evidence underlying a pre-specified mechanism (for example, “What is the evidence that milk causes prostate cancer by up-regulating the insulin-like growth factor (IGF) pathway?”).*

#### Defining the research question

Researchers should specify the modifiable exposure and the outcomes of interest.

**Notes**

The exposure may be a behavioural, dietary or lifestyle (e.g. physical activity) factor, it may be an environmental factor (air pollution), or it may be a phenotype such as obesity or the metabolic syndrome. The outcome can be any phenotype or disease of interest, for example it may be incidence of a specific cancer (or several cancers), or cancer progression, or another disease phenotype such as cardiovascular disease.

#### Conducting the literature search

##### Develop lists of search terms, define which databases to search and carry-out searches.

**Notes**

A framework for comprehensively identifying the search terms for Stage 1 searches is provided in Supplementary Table 1. This framework lists those factors to consider when developing the search terms for the project-specific exposures, intermediate phenotypes and endpoints. To develop a list of intermediate phenotypes, we recommend considering the biological processes which may lead to the disease of interest. This could be achieved by referring to some important reviews on potential mechanisms to the outcome of interest.

The search terms for Stage 1 should be PubMed/MEDLINE MeSH terms. It may be necessary to consult with a librarian or information specialist on the search terms to be used to ensure that appropriate synonyms and spellings are included. In addition MeSH on demand (<https://www.nlm.nih.gov/mesh/MeSHonDemand.html>) can be used to identify MeSH terms relevant to your search terms of interest.

*Combining search terms:* To identify relevant studies, the selected search terms may be combined as follows:

(exposure terms) AND (intermediate phenotype terms) OR

(intermediate phenotype terms) AND (outcome terms) OR

(exposure terms) AND (outcome terms)

The relevant studies are those reporting on associations between exposure and outcome, biological mechanism and outcome and between exposure and biological mechanism. Studies linking all three concepts are likely to be particularly relevant but are less common.

PubMed/MEDLINE is the best known and most widely used biomedical database, including health care, biomedical sciences and clinical sciences and will give a good overview of the literature on mechanisms relating to disease. As Stage 1 is essentially a scoping exercise and not systematic we recommend just using PubMed/MEDLINE for these searches.

#### Analysing the findings from the search

The purpose of this step is to identify potential mechanisms and to summarize the quantity of evidence on each mechanism

**Notes**

The above searches are likely to uncover a large number of potentially relevant studies. We have, therefore, devised an automated process (‘Text Mining for Mechanism Prioritisation’, TeMMPo). This tool can be accessed at <https://www.temmpo.org.uk/>. The programme allows users to upload the results of their MEDLINE or PubMed searches which are then displayed according to the intermediate phenotypes included in the study.

**Use of TeMMPo (‘Text Mining for Mechanism Prioritisation’** [**https://www.temmpo.org.uk/**](https://www.temmpo.org.uk/)**)** **for mechanism prioritization**

TeMMPo is an internet based tool which allows quantification and visualisation of the evidence underlying each step in the mechanistic pathway.

The quantification algorithm requires the user to input three lists of MeSH terms: exposures, intermediate phenotypes and outcomes. It uses export files from the electronic bibliographic databases used for the stage 1 searches, which need to be uploaded as the source data. The citations are then read into memory as a list of “citation” objects to identify the co-occurrence of at least two of: exposure, intermediate phenotype and exposure in each citation. At the end of the process, the quantity of evidence linking exposure to intermediate phenotype and intermediate phenotype to outcome (E → IP, IP → O, E → O) is assembled for graphical presentation and as a simple text table. Biological mechanisms that are not connected with at least one exposure and one outcome are not included in the counts. The most frequently occurring intermediate phenotypes are then included in the list of biological mechanisms for counting. Both approaches are coded in the Python programming language and require text input (standard MEDLINE text export) and comprehensive candidate lists of exposures, biological mechanisms and outcomes (as developed in Stage 1,Step 2).

For each biological mechanism, a summary table of counts of abstracts which mention exposure-biological mechanism and biological mechanism-outcome relationships, and a weighted average of the quantity of studies for these two groups is generated, along with a graphical representation (a Sankey plot) where weights (indicated by the thicknesses of the connecting lines [edges]) represent the quantity of studies. Figure 2 shows a Sankey plot indicating the quantity of studies linking milk with an intermediate phenotype and the quantity of studies linking the same intermediate phenotypes with a prostate cancer outcome. The automated approach makes two major assumptions:

* that the co-occurrence of a biological mechanism with exposure or outcome in the literature represents an association
* that the mechanisms are represented by a single mediating factor.

In addition, novel pathways will be underrepresented in this approach as they are likely to have fewer studies.

This approach does not address issues of study type, quality, direction and significance of results.

### Stage 2- systematic review of the evidence linking an exposure with intermediate phenotypes in the mechanistic pathway of interest, and evidence linking the same intermediate phenotypes with the outcome.

Having identified potential mechanisms underlying a particular exposure-outcome association, Stage 2 addresses an objective to systematically review the evidence underlying one or more specific mechanisms.

#### Stage 2, Step 1 Specify the research objectives

A detailed research question will help to guide the search strategy and will lead to a more defined and systematic search.

**Notes**

By analogy with the established ‘PICO’ framework for considering research questions on the effects of interventions (participant(s); intervention(s); comparator(s); outcome(s)), in the current context the research question can be formulated by specifying the population(s), exposure(s), mechanism(s) and outcome(s) of interest (‘PEMO’). Where an intermediate phenotype has been identified, there is likely to be one PEMO which addresses the exposure-intermediate phenotype relationship (E → IP) and a separate one which links the intermediate phenotype and outcome (IP → O). Constructing a causal diagram depicting the pathway from exposure to outcome may help to focus the review questions; this could include the interplay between multiple intermediates where this information is already known.

Relevant studies will be those that link the exposure of interest (e.g. milk) to a biological mechanism (e.g.IGFs), as well as studies which link the same biological mechanism to the cancer outcome of interest (e.g. prostate cancer).

The PEMOs will also indicate project-specific inclusion and exclusion criteria that will need to be applied by the reviewer.

#### Stage 2, Step 2 Searching for studies

Targeted searches aimed at synthesising the evidence underlying specific mechanistic pathways of interest should be carried out.

**Notes**

Searches will be similar to searches carried out in conventional systematic reviews, except that researchers should carry out two sets of searches, one to identify studies which address the exposure-intermediate phenotype (mechanism) relationship (E → IP) and secondly to identify studies which investigate the link between the intermediate phenotype and outcome (IP → O). The following databases can be searched in Stage 2:

* MEDLINE (1950-)
* EMBASE (1980-)
* CINAHL (1981-)
* The institute for Scientific Information (ISI) Web of Knowledge
* Google scholar
* AMED (Allied and Complimentary Medicine Database)
* Bibliographic databases for grey literature, including theses, conference abstracts and book chapters.

We recommend that no language or study design limits are placed on the search. Any other limits, such as coverage dates, should be pre-specified in the project specific protocol.

#### Stage 2, Step 3 Inclusion and exclusion of studies

The titles and abstracts should initially be screened, followed by full text review if necessary and predefined inclusion/exclusion applied.

**Notes**

Screening should be carried out by at least two reviewers. Disagreements should be resolved by reaching consensus through discussion. A review of the reference list of included studies can be undertaken to identify additional references that were not picked up in the electronic searches. A study flow diagram based on the PRISMA statement1 should be included to show the number of records identified in the literature searches, the number of studies included and excluded, and the reasons for exclusions.

#### Stage 2, Step 4 Data extraction

We recommend that at least two reviewers should extract the data according to a pre-specified list of variables and resolve disagreements by reaching consensus through discussion.

**Notes**

Data should be extracted for mechanistic studies. Reviewers should devise their own list of variables to extract. For each study, baseline characteristics should be tabulated (e.g., study size, population, intervention or exposure, comparator, outcomes, follow-up period) and the citation provided. We recommend extracting the data listed by study type in Supplementary Table 2 in order to assess study quality.

#### Stage 2, Step 5 Assessment of study quality

Study quality should be assessed using validated risk of bias tools.

**Notes**

Specific tools are available for randomized controlled trials2 observational studies3 and for animal studies we recommend the SYRCLE (Systematic Review Centre for Laboratory animal Experimentation) tool,4 Risk of bias should be assessed independently by at least two reviewers and discrepancies resolved by discussion to reach consensus. A risk of bias table should be completed for each individual study included in the review**.**

##### *Systematic reviews*

Where an up-to-date systematic review is available, we would recommend using the data from this instead of extracting data, assessing study quality and analysing data separately for each study in the systematic review, providing the review is of high quality. The quality of the review can be assessed using the ROBIS tool which has been developed to assess the risk of bias in systematic reviews.5   If a review is judged to be applicable and of high quality (ie low risk of bias), the information in the review can be used, although it may be appropriate to update the literature searches if the review was not undertaken recently.

##### *Cell line (in-vitro) studies*

The following points could be considered when assessing the quality of *in vitro* or cell culture-based evidence:

* Have the cells been obtained from a validated repository that guarantees cell verification or have the cells been appropriately independently verified? If not the data from such studies may still be useful but perhaps less relevant.
* Have the experiments been repeated a sufficient number of times and were appropriate controls included?
* Are culture conditions comparable between different studies?
* How many different cell lines from the same cancer type were 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.
* Were cell lines from different cancer types compared? This implies an important effect that is relevant more generally to cancer cells.
* Selective reporting: are only selected results from several cell line experiments reported?

#### Stage 2, Step 6 Synthesis of data from individual studies

For each study, baseline characteristics, results and data on risk of bias should be presented. In addition, where appropriate results from similar studies may be combined.

**Notes**

For the outcomes considered it is important to present, for each study: (a) simple summary data for each intervention or exposure group; and (b) effect estimates, confidence intervals and p-values.

We recommend the use of formal meta-analysis of comparable studies where possible and appropriate, but we anticipate that many mechanistic studies will be too heterogeneous (in terms of exposure and outcome definitions) to combine and therefore it is likely that some studies will only be amenable to a narrative summary of the results. Meta-analyses may follow conventional methods as used for randomised trials6 or may use more sophisticated methods such as have been proposed to combine studies in humans and animals.7 Potential sources of heterogeneity should be examined in pre-specified sub-group analyses and, if there are sufficient studies, by meta-regression.8 When meta-analysis is not appropriate, other methods may be used to graphically represent the datasuch as Albatross plots,9 without the underlying assumptions of homogeneity within the studies10.

**Investigating heterogeneity**

Investigators should pre-specify variables potentially contributing to heterogeneity in the study, based on biological knowledge (e.g. age; sex; duration of follow-up; specific clinical groups) and investigate heterogeneity by methodological variables, including study design and responses to the signalling questions for each of the risk of bias domains.

**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 Begg11 and Egger12 tests can be used to examine for association between effect sizes and study sizes (essentially sample size), and such an association may reflect publication bias. These approaches may not be possible because of a lack of a sufficient number of similar studies with the same exposures and outcomes measured.

Ioannidis and Trikalinos13 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, common type of interventions, common type of subjects, common methodology, common research environments, common 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 amongst 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 grey literature and/or contacting researchers working in the field) to determine whether relevant unpublished experiments or observational studies have been carried out. It can be difficult to be 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 grey literature studies with those which have been published to determine if 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.

#### Stage 2, Step 7 Assessing the strength of the overall body of evidence within evidence streams

Once the synthesis of evidence has been completed, researchers should assess the body of evidence within evidence streams (i.e. separately for human and animal studies which closely mimic human disease). At this point it will be important to determine the relevance of the studies to the human disease process (Supplementary Box 1).

**Notes**

**Application of the GRADE framework**

The body of evidence should be assessed according to the GRADE framework,14 which has been adopted by many organisations around the world, including the Cochrane Collaboration. GRADE provides a broad interpretation of the quality of a body of evidence using a systematic approach, making judgements explicit and transparent. It includes considerations of the following five key domains, which we address in turn:

**1)Risk of bias**

Risk of bias is assessed both within and across studies. Across studies, the assessment should consider the proportion of evidence with different risks of bias (i.e. if the majority of studies are of low risk of bias, the overall rating would be low even if some studies were at a high risk of bias). Where studies are of different sizes, this should be taken into account by giving larger studies greater weighting in the overall assessment rather than simply by counting the number of studies at each level of risk of bias.

**2) Indirectness**

Indirectness (relevance) refers to the applicability of the results to the population; interventions and exposures (e.g. the route of exposure, whether the dose reflects a typical human exposure, whether the exposure occurs in the appropriate biological window to affect the outcome); and outcomes (e.g. use of surrogate outcomes, outcomes assessed an adequate amount of time after exposure). Indirectness is assessed both within and across studies.

**3) Inconsistency**

Inconsistency across studies is based on point estimates varying widely, confidence intervals showing minimal or no overlap and large I2 values. For animal models it will also be important to assess consistency in the light of the fact that studies may have used different animal models and perform slightly different experiments.

**4) Imprecision**

Confidence intervals capture the extent of imprecision. The body of evidence will be considered imprecise if made up of a small number of studies, which all have wide confidence intervals.

**5) Publication bias**

Issues to consider when assessing publication bias are discussed above. Note that the GRADE domain specifically pertains to non-availability of information about whole studies, since selective reporting within a study is covered in risk-of-bias criteria. When serious concern exists about publication bias, this will significantly decrease confidence in the body of evidence.

**Overall GRADE assessment**

The overall quality of the body of evidence is summarised within evidence streams (i.e. separately for human and animal studies) using four levels:

* High quality (further research is very unlikely to change our confidence in the estimate of effect)
* Moderate quality (further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate)
* Low quality (further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate)
* Very low quality (any estimate of effect is very uncertain)

The findings are tabulated in a GRADE table using the following algorithm:

1. A starting rating is given based on the design of the individual studies. Specifically, a starting rating of ‘High quality’ is given if there is evidence from human or animal RCTs, but a starting rating of ‘Low quality’ is given where the only evidence is from human or animal observational studies.
2. Then five reasons for downgrading the overall quality of the body of evidence are considered according to the five domains listed above: risk of bias; unexplained inconsistency; indirectness (relevance); imprecision and publication bias, with downward adjustments to the starting rating if serious concerns are noted within these domains.
3. In addition, three reasons are available for upgrading the overall quality of the body of evidence: strength of association (a large magnitude of effect that is unlikely to be explained by bias or confounding); a clear dose-response; and minimal impact of bias and confounding. Again, adjustments are made to the rating if any of these are found.

Notes should be included in the table to explain decisions and acknowledge borderline decisions. Different study designs (e.g. RCT and observational studies) may be graded separately in different tables or in the same table, but conclusions should be based on the evidence with the highest confidence. In the summary of findings section there is a qualitative assessment of confidence (not a score). More details on how to implement the GRADE approach are provided at: <http://www.gradeworkinggroup.org/index.htm>. The judgments involved in GRADE evaluation are inherently subjective; however, this process provides a transparent framework to document and justify the decisions made to classify the overall quality of evidence.

Questions for assessing the indirectness of animal studies to human disease are provided in Supplementary Box 2.

We do not recommend subjecting cell line or *in vitro* studies or animal studies which do not closely reflect the human disease process to a GRADE assessment (Supplementary Box. The role of such studies in the integration of evidence across evidence streams is described in Stage 2, Step 8, below.

#### Stage 2, Step 8 Assessing the strength of evidence across evidence streams

Once the overall body of evidence has been assessed separately for animal and human studies using GRADE, we recommend integrating the highest levels of evidence from both evidence streams using the simple 3 by 3 schema shown in Figure 5 (adapted from15)

**Notes**

To reach an overall conclusion based on the strength of evidence for the particular mechanism under investigation. There are four conclusion categories:

* Strong evidence in support of the underlying mechanism
* Moderate evidence in support of the underlying mechanism
* Weak evidence in support of the underlying mechanism
* Inconclusive evidence in support of the underlying mechanism

#### Because of the inherent difficulty in proving a negative, a conclusion of evidence that the mechanism under study is not involved in the pathway between exposure and outcome is only reached when the body of evidence is of high quality. A low or moderate quality of evidence results in a conclusion of inadequate evidence to reach a conclusion that the mechanism is not involved.

#### Stage 2, Step 9 Synthesis of cell line and animal studies which do not closely mimic the human cancer process or which have cancer-related outcomes (i.e. hallmarks of cancer).

# A summary table of cell studies and a separate table of animal studies should be presented, which outlines the experiments performed and the main findings. In addition reviewers may find it helpful to present the findings from high quality cell line or animal studies in the form of a causal diagram.

# Notes

Cell line and those animal studies which are not deemed to closely reflect the human cancer process (e.g. xenograft models) are also considered in relation to the final conclusions. Such data may provide strong support for the biological plausibility of the mechanism or may provide evidence that the mechanism is not biologically plausible. These studies should be assessed for quality, as described in Stage 2, Step 5. Reviewers can then decide to present information from all the cell line or animal studies or for high quality studies only, according to predefined criteria (e.g. only cell studies where cell lines have been verified). A conclusion should be reached on whether the studies support the biological plausibility of the causal pathway which is being investigated.

**References**

1. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. PLoS Medicine 2009; 6(6): e100097.

2. Higgins JPT, Altman DG (eds) 2008. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.0.1 (updated September 2008). The Cochrane Collaboration. Available from [www.cochrane-handbook.org](http://www.cochrane-handbook.org).

3. Sterne J, Higgins J, Reeves B. Extending the risk of bias tool to allow for assessment of non-randomised studies, cluster-randomised trials and cross-over trials: a Cochrane methods innovation fund project (Workshop). In: The Cochrane Collaboration. 21st Cochrane Colloquium Abstract Book, 2013, pp 203-204. Available from: <http://2013.colloquium.cochrane.org/abstract-book>

4. Hooijmans CR, Rovers MM, de Vries RBM, Leenaars M, Ritskes-Hoitinga, M, Langendam, MW. SYRCLE´s risk of bias tool for animal studies. BMC Medical Research Methodology 2014; 14:43: doi: 10.1186/1471-2288-14-43.

5. Whiting P, Savović J, Higgins JPT, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R, the ROBIS group. ROBIS: a new tool to assess the risk of bias in systematic reviews. *Journal of Clinical Epidemiology* 2015. Published online 16 Jun 2015; doi: 10.1016/j.jclinepi.2015.06.005

6. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled Clinical Trials 1986; 7: 177-188.

7. Peters, J. L., Rushton, L., Sutton, A. J., Jones, D. R., Abrams, K. R. and Mugglestone, M. A. (2005), Bayesian methods for the cross-design synthesis of epidemiological and toxicological evidence. Journal of the Royal Statistical Society: Series C (Applied Statistics), 54: 159–172

8. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Statistics in Medicine 1999; 18: 2693-2708.

9. 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. *Research Synthesis Methods.*2017. https://doi.org/10.1002/jrsm.1239

10. Ogilvie D, Fayter D, Petticrew M, Sowden A, Thomas S, Whitehead M, Worthy G [The harvest plot: a method for synthesising evidence about the differential effects of interventions.](http://www.ncbi.nlm.nih.gov/pubmed/18298827) BMC Med Res Methodol. 2008 Feb 25;8:8. doi: 10.1186/1471-2288-8-8. Review.

11. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994; 50: 1088-101.

12. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ 1997; 315: 629-34

13. [Ioannidis JP](http://www.ncbi.nlm.nih.gov/pubmed?term=Ioannidis%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=17715249)A, [Trikalinos TA](http://www.ncbi.nlm.nih.gov/pubmed?term=Trikalinos%20TA%5BAuthor%5D&cauthor=true&cauthor_uid=17715249). An exploratory test for an excess of significant findings. [Clin Trials.](http://www.ncbi.nlm.nih.gov/pubmed/17715249) 2007;4:245-53.

14Guyatt 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: 924-926.

154. 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.

##### Supplementary Box 1 - Classification of studies according to relevance to humans

We recommend a distinction between two categories of studies:

1. Studies in humans, and animal models that closely mimic human disease
2. Studies in animal models that do not closely mimic human disease, and cell line studies.

To distinguish between animal models of the two types we suggest applying the question “Has the disease arisen in the animal model?” For cancer this is easy to distinguish and is because transplantable models represent cancers which are already highly evolved as they have grown well in culture and are typically a very aggressive biological phenotype; as such they do not closely mimic human cancers and are very unlikely to give useful information about the aetiology of cancer risk or progression.

Studies that do not closely mimic human diseases should not be used to determine the strength of the evidence underlying a particular mechanistic pathway in humans. Instead they can be assessed alongside cell line studies to determine whether they provide evidence for the biological plausibility of the proposed mechanism (Figure 5, Step 8).

**Supplementary Box 2- Criteria for assessing the indirectness of animal studies when applying the GRADE framework**

To assess indirectness in animal studies we recommend using the following screening questions. 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:

* Is the exposure applied via a route which is comparable with that in humans, and a mode which addresses the research question? *(e.g. If the interest is in a food exposure, then this should be ingested in the animal model, but if it is a component such as a particular protein it may be appropriate to inject this into the animal)*
* Is the level and frequency of exposure comparable with that which humans may experience after taking into account species differences in pharmacokinetics and pharmacodynamics, or is the dose justified within the study? *(doses which are much greater than would be tolerated by humans are unlikely to reflect exposures which are seen in humans)*
* Is the cancer induced (i.e. by a virus, radiation, chemical agent or transgenically)? (*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 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 a much longer period of follow-up to produce relevant data*.
* Does the study explore mechanisms or pathways of cancer development?
* Is the outcome being assessed cancer incidence or progression rather than surrogate measures of tumour activity such as tumour size or number of tumours?
* Does the outcome being assessed mimic outcomes found in humans? More specifically, does the tumour mimic the human disease in terms of the organ or tissue affected, and at histopathological level (tissue patterns, or cell surface or intracellular protein expression levels) or the 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)?

|  |  |
| --- | --- |
| Domain | Considerations |
| Exposure | What are the constituents of the exposure?What foods is the exposure contained in?What are the different sources of exposure (e.g. dietary vitamin D vs sunlight)?What are the potential confounders?How is the exposure produced, processed and packaged?Are there geographical or temporal variations in exposure?Do contaminants need to be considered  |
| Intermediate phenotypes | What are the relevant mechanistic targets?Are there specific physiological process relevant to the disease of interest? (e.g. gastric reflux in oesophageal cancer; insulin sensitivity in Type 2 diabetes)  |
| Outcome | Is the interest in risk of disease or in disease progression?Are specific subtypes of interest? |

Supplementary Table 1 Framework for developing lists of search terms for exposures, intermediate phenotypes and outcome

Human experimental studies

* Study name
* Study population (location, sample size, year of recruitment, ethnicity, age and gender distributions)
* Experimental design
* Whether the intervention was randomized and, if so details of randomization
* Whether allocation concealment was carried out
* Whether participants, and/or healthcare professionals, and outcome assessors were blinded to randomization
* Details of the control/intervention group
* Effect estimates comparing intervention versus control groups (experimental studies) with corresponding 95% confidence intervals (CIs)
* Duration of follow-up
* Intermediate phenotypes analysed
* Cancer outcomes analysed
* Details of confounders adjusted for or matched on, and effect estimates for both crude and fully adjusted models
* Subgroup analyses
* Data to assess the risk of bias (see Stage 2, Step 5**)**

Human observational studies

* Study design (e.g. case-control, cohort etc. and whether prospective or retrospective, population-based or hospital-based)
* Study name
* Study population (location, sample size, year of recruitment, ethnicity, age and gender distributions)
* Details of the exposures measured and methods of measurement (observational studies)
* Duration of follow-up
* Intermediate phenotypes analysed
* Cancer outcomes analysed
* How the exposure levels were presented (quantiles/categories, continuous or mean differences), and how the results were analysed
* Which group (ie total population, cases or controls) was used to calculated quantiles if quantiles were used
* The association in exposed versus unexposed groups with corresponding 95% confidence intervals (CIs)
* Details of confounders adjusted for or matched on, and effect estimates for both crude and fully adjusted models
* Subgroup analyses
* Details of other potential confounders where reported Data to assess the risk of bias (see Stage 2, Step 5**)**

Animal studies

* Number and type of animal model/mod**e**ls used
* Animal housing conditions for experimental and control group
* Nutrition intake/exercise levels for experimental and control group
* Experimental design (ie whether randomized and if so how)
* Duration of follow-up
* Intermediate phenotypes analysed
* Cancer outcomes analysed
* Whether animal handler and/or outcome assessor was blinded to intervention
* Details of confounders adjusted for or matched on, and effect estimates for both crude and fully adjusted models
* Subgroup analyses
* Effect estimates comparing intervention versus control groups (experimental studies) with corresponding 95% confidence intervals (CIs)
* Details of statistical methods
* Data to assess the risk of bias (see Stage 2, Step 5**)**

Cell line studies

Names of cell lines

Whether cell lines were established patient-derived tumour cells lines or freshly isolated primary cells

Whether cell lines were authenticated

Culture conditions (e.g. concentrations of glucose in the growth media)

Whether cells were grown in 3D rather than 2D

The treatment regime

Details of laboratory procedures (e.g. cell proliferation methods)

Details of statistical methods and mean levels of proliferation (and corresponding standard deviations, standard errors or 95% confidence interval).

**Supplementary Table 2-List of recommended variables to extract by study type**