The Grading Criteria with the Global Cancer Update Programme

VERSION DATE: 10-NOVEMBER-2023

Table of Contents

BACKGROUND

Cancer is an important and increasing problem: Globocan estimates that the number of new cases of cancer, worldwide, will increase from 19.3 million per year in 2020 to 30.2 million in 2040 and that mortality from the disease will increase from 9.96 million to 16.3 million in the same time period. Diet, nutrition, physical activity, and body weight are now recognised as factors in its causation and progression, and it is important to identify the links between them as a basis for clinical and public health policy.

These criteria provide the framework for the Panel's judgements on the evidence reviews to allow them to draw conclusions, both for cancer incidence (to underpin recommendations or guidance for cancer prevention), and for cancer survivors (to underpin recommendations or guidance for interventions to reduce the risk of adverse cancer related outcomes). The purpose of the criteria is to provide a standardised framework within which the Panel can, as objectively as possible, categorise the likelihood that a dietary, nutritional or physical activity exposure is either causally related to risk of cancer (for cancer incidence and prevention), or has an effect on a cancer related outcome after a cancer diagnosis (for cancer survivors). See below further information about how we are treating incidence and survival evidence, and the terminology used.

Cancer incidence criteria

For cancer incidence, the criteria are based on the framework developed by Bradford Hill, to identify from observational evidence the likelihood of a specific exposure being a cause of a specific cancer. Although in theory it might appear desirable to base such decisions on randomised controlled trials (RCTs), for a variety of reasons detailed elsewhere (e.g. [1]), such trials may be impractical or flawed. Although trials can contribute to the inference of causation as part of an overall body of evidence, that also includes observational and mechanistic studies, they are not necessarily regarded as a higher form of evidence.

Cancer survivors criteria

Research with cancer survivors involves different scientific questions and deals with different data to cancer incidence. The GRADE criteria are more appropriate for the survivorship setting since the research addresses the certainty of an effect as opposed to population cancer prevention which addresses the likelihood of causality.
The criteria for cancer survivors are based on (though not absolutely identical to) the GRADE framework, which is designed to evaluate the certainty that an intervention will lead to an effect. This can be based on direct evidence from RCTs, or inferred from observational evidence, supported by mechanistic data. In particular the process of moving from evidence judgements to recommendations differs slightly from GRADE, though it is based on the same principles.

Terminology for grading the evidence

The terminology for strong evidence of likely causality for cancer incidence (Convincing and Probable) has been used in all three WCRF/AICR Expert Reports since 1997. The use of Limited (either suggestive or no conclusion) for weak evidence of causality has been used since 2007 in the Second and Third Expert Reports. These terms are widely recognised and have proved useful in practice (see page 5 for a full explanation).

The terminology for cancer survivors is based on the GRADE categories of High, Moderate, Low, or Very Low certainty of an effect (see page 8 for a full explanation)

The different terminology between the two contexts emphasises the different scientific questions being addressed, as well as the different evidence needs, and the weight given to various types of evidence. In addition we have criteria for an additional level of evidence "Substantial effect unlikely".

Using the criteria to make recommendations and develop guidance

In both contexts, strong evidence would be considered for developing recommendations, where implementation would be expected to lead to the desired outcome, while weak evidence might be used to provide less rigid guidance, where the confidence of achieving the desired outcome would be less.

These criteria, for both cancer incidence and survivors reviews, comprise a description of the minimum evidence needs, with both upgrading and downgrading factors (including definitions of "well-designed" studies), to achieve a particular level of conclusion.

Information about how mechanistic evidence supports the grading criteria, additional context, and information about how the evidence gradings appear in evidence matrices for the CUP Global review outputs is also presented.

For cancer incidence and prevention, the purpose of the criteria is to help the Panel distinguish between true causal links from associations that are not judged to be causal. Reducing exposure to a factor that is causally linked to increased cancer risk, or increasing exposure to a protective factor would be expected to lower cancer incidence in populations, and to lower cancer risk in individuals in that population. Changing the level of exposure may refer to the number of individuals exposed or to the intensity of the exposure. In addition, while ascertainment of the level of an exposure may have been done on only one occasion, this may reflect exposure over a prolonged period. Influencing risk by changing exposure may require that change to be effected over an equally prolonged period, often decades.

For cancer survivors, the purpose of the criteria is to judge the degree of certainty that a particular dietary, nutritional, body weight or physical activity intervention in a population defined by a specific cancer diagnosis, would lead to a change in defined cancer related outcomes.

Both for cancer incidence and survivors, the important distinction is between evidence that is judged strong enough to make a recommendation, and evidence that is not strong

enough to justify a recommendation. The highest two categories of evidence, (for cancer incidence, convincing or probable; for cancer survivors, high or moderate certainty are regarded as strong enough to support recommendations. Evidence judged to fall into the limited/ or low certainty categories is generally not regarded as strong enough to support recommendations. Such weaker evidence might nevertheless offer pragmatic guidance, but where there is less confidence that implementation would lead to the desired outcome. Guidance would only be developed where there is deemed a low likelihood of harm.

Importantly the criteria allow for flexibility through specified upgrading or downgrading factors – characteristics of the evidence that tend to strengthen or weaken confidence in a conclusion, such as direct evidence from RCTs, a large and unbiased effect size or robust human experimental evidence.

The Panel has sometimes not made recommendations despite strong evidence. This might be because of potentially adverse effects on one cancer, despite evidence of protection for another cancer or disease (for example, calcium and/or dairy with risk of prostate and colorectal cancer); or because it is not possible to craft a recommendation that is useful in practice (as is the case for adult attained height). For further information see: Recommendations and public health and policy implications. Conversely, the Panel may sometimes make a firm recommendation when there is a large volume of consistent evidence for subgroups. For example, non-starchy vegetables and fruit (as a group) and aerodigestive cancers (as a group) was judged as probable decreases risk, although the evidence was only judged to be suggestive for each subgroup (see Recommendations and public health and policy implications). The Panel also considered the implications for other diseases when making recommendations, and so the recommendations can be considered to reduce the risk of other non-communicable diseases related to diet, nutrition, body weight, and physical activity.

The overarching intention of CUP Global is to develop recommendations and guidance that are applicable to the general population, as well as specifically for people living with and beyond cancer. In addition, CUP Global aims to refine the existing recommendation to be specific for populations defined, for example by age. This work will be developed via several research areas including lifecourse research – which seeks to better understand how early life exposures impact cancer risk in adult life – and other specific reviews (for example, reviews of the evidence related to survival following a diagnosis of cancer). The information gathered within this work might then enable the development of population-specific recommendations. Recommendations for cancer survivors will be cancer-specific and will be based on the evidence gradings from reviews on survival outcomes as well as patient reported outcome measures (e.g. quality of life, fatigue). Where evidence is in not sufficiently strong to develop recommendations, consideration will be given to whether guidance can be developed. This will be developed using the evidence gradings along with expert opinion and user involvement. See the Appendix for an overview of the CUP Global process for developing guidance.

There may be occasions where the Panel will consider all the evidence on a specific topic and based on this evidence may wish to make an overarching conclusion statement. This considers the totality of the available evidence alongside expert opinion.

CUP Global recognises the importance of health inequalities and broader issues of diversity, equity and inclusion. Evidence from diverse global locations and populations might, in the future, be used in drawing conclusions or making recommendations/guidance.

CRITERIA FOR GRADING EVIDENCE FOR CANCER PREVENTION (INCIDENCE)

The criteria were developed as a means of operationalising, within the context of diet, nutrition, physical activity and cancer, the factors identified by Bradford Hill [2] as contributing to an inference of causation from epidemiological evidence. These criteria were developed by an independent expert Methodology Task Force, revised by the Expert Panel for the 2007 Second Expert Report (Chapter 3 of the 2007 Second Expert Report [3]), and were also used for the Third Expert Report [3].

The current criteria were reviewed and further revised to improve clarity and utility (but not substantively changed) by the previous CUP Panel, the CUP Transition Panel, and the relevant Expert Committees in 2022 for CUP Global. The criteria lead to five possible levels of conclusion. In effect, the criteria define these terms, which are those used in the matrices:

1. Convincingly causal
2. Probably causal
3. Limited evidence but suggestive of a possible causal relationship
4. Limited evidence and no conclusion of a causal relationship possible;
5. Substantial effect on risk unlikely (effectively a 'no effect' conclusion, though it is never possible to completely exclude a small effect).

Numbers 1, 2 and 5 are considered strong evidence.

Convincing (Strong evidence)

Evidence considered strong enough to support a judgement of a convincingly causal (or protective) relationship justifies considering making a corresponding recommendation designed to reduce the risk of cancer. The evidence is robust enough to be unlikely to change in the foreseeable future as new evidence accumulates.

All of the following are generally required:

- Evidence of a significant association from more than one study type (RCTs, cohort, or pooled analyses of individual data from these studies)
- Evidence from at least two independent cohort studies (where there are at least two studies that show significant associations or a significant meta-analysis)
- No substantial unexplained heterogeneity within or between study types or in different populations relating to the presence or absence of an association, or direction of effect. In short, consistent findings across multiple studies
- No substantial evidence of publication bias

- Good-quality studies to exclude (with confidence) the possibility that the observed association results from random or systematic error, including confounding, measurement error and selection bias. The features of well designed studies are described in more detail below (see page 13).
- Presence of a plausible biological gradient ('dose–response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.
- Strong and plausible experimental evidence, either from human studies or relevant animal models, that typical human exposures can lead to relevant cancer outcomes.

Probable (strong evidence)

Evidence strong enough to support a judgement of a probable causal (or protective) relationship, which generally justifies recommendations designed to reduce the risk of cancer.

All of the following are generally required:

- Evidence from at least two independent cohort studies (where there are at least two studies that show significant associations or a significant meta-analysis).
- No substantial unexplained heterogeneity between or within study types in the presence or absence of an association, or direction of effect.
- No substantial evidence of publication bias
- Good-quality studies to exclude with confidence the possibility that the observed association results from random or systematic error, including confounding, measurement error and selection bias.
- Evidence for biological plausibility.

Substantial effect on risk unlikely (strong evidence)

Evidence is strong enough to support a judgement that a particular food/diet, nutrition, bodyweight or physical activity exposure is unlikely to have a substantial causal relation to a cancer outcome. The evidence should be robust enough to be unlikely to change in the foreseeable future as new evidence accumulates.

All of the following are generally required:

- Evidence from more than one study type (RCTs, cohort or pooled analyses of individual data of these studies).
- Evidence from at least two independent cohort studies (where there are at least two studies that show significant associations or a significant meta-analysis).
- Summary estimate of effect close to 1.0 for comparison of high- versus low-exposure categories.
- No substantial unexplained heterogeneity within or between study types or in different populations.
- Good-quality studies to exclude, with confidence, the possibility that the absence of an observed association results from random or systematic error, including

inadequate power, imprecision or error in exposure measurement, inadequate range of exposure, confounding and selection bias.
- Absence of a demonstrable biological gradient ('dose–response'). Absence of strong and plausible experimental evidence, from either human studies or relevant animal models, that typical human exposure levels lead to relevant cancer outcomes.

Factors that might misleadingly imply an absence of effect include imprecision of the exposure assessment, insufficient range of exposure in the study population, and inadequate statistical power. Methodological issues such as these, and in other study design attributes, might lead to a false conclusion of no effect.

The presence of a plausible, relevant biological mechanism does not necessarily rule out a judgement of 'substantial effect on risk unlikely'. But the presence of robust evidence from appropriate animal models or human studies, that a specific mechanism exists or that typical exposures can lead to cancer outcomes, argues against such a judgement.

Because of the uncertainty inherent in concluding that an exposure has no effect on risk, the criteria used to judge an exposure 'substantial effect on risk unlikely' are roughly equivalent to the criteria used with at least a 'probable' level of confidence. Conclusions of 'substantial effect on risk unlikely' with a lower confidence than this would not be helpful and could overlap with judgements of 'limited – suggestive' or 'limited – no conclusion'.

Limited – suggestive

Evidence that is too limited to permit a probable or convincing causal judgement but is suggestive of a direction of effect. The evidence may be limited in amount or by methodological flaws but shows a generally consistent direction of effect. This judgement is broad and includes associations where the evidence falls only slightly below that required to infer a probably causal association through to those where the evidence is only marginally strong enough to identify a direction of effect. This judgement is very rarely sufficient to justify recommendations designed to reduce the risk of cancer. Any exceptions to this require a special and explicit justification.

All of the following are generally required:
- Evidence from at least two independent cohort studies (where there are at least two studies that show significant associations or a significant meta-analysis)
- The direction of effect is generally consistent though some unexplained heterogeneity may be present.
- There may be moderate or high risk of bias in some studies
- There may be moderate or high risk of publication bias
- Evidence of biological plausibility.

Limited – no conclusion

Evidence is so limited that no firm conclusion can be made. This judgement represents an entry level and is intended to allow any exposure for which there are sufficient data to warrant Panel consideration, but where insufficient evidence exists to permit a more

definitive grading. This does not necessarily mean a limited quantity of evidence. A body of evidence for a particular exposure might be graded 'limited – no conclusion' for a number of reasons. The evidence may be limited by the amount of evidence in terms of the number of studies available, by inconsistency of direction of effect, by methodological flaws (for example, lack of adjustment for known confounders) or by any combination of these factors.

There is a difference between a 'Limited – no conclusion' grading and one that is 'Substantial effect on risk unlikely (strong evidence). When an exposure is graded 'limited – no conclusion', this does not necessarily indicate that the Panel has judged that there is evidence of no relationship. With further good-quality research, any exposure graded in this way might, in the future, be shown to increase or decrease the risk of cancer. Where there is sufficient evidence to give confidence that an exposure is unlikely to have an effect on cancer risk, this exposure will be judged 'substantial effect on risk unlikely'.

There are also many exposures for which there is such limited evidence that no judgement is possible. In these cases, evidence is recorded on the World Cancer Research Fund International website (dietandcancerreport.org). However, such evidence is usually not included in the summaries. The systematic literature reviews undertaken previously within CUP will remain available on the WCRF website, while those produced as part of CUP Global will be added to the website as they become available. When within our control, WCRF will make peer-reviewed papers, produced as part of CUP Global, open access.

CRITERIA FOR GRADING EVIDENCE FOR CANCER SURVIVORS

These criteria have been developed based on the GRADE framework as a tool to judge the degree of certainty that a specific intervention, in people with a specific cancer diagnosis, will affect the risk of developing a specific cancer-related outcome. We will work with the relevant Expert Committees, as well as the Panel, to progress the cancer survivors criteria. Once finalised, these criteria will differ slightly from the GRADE process. Firstly, there is an additional level of evidence ("Substantial effect unlikely"). Secondly, the criteria include a consideration of mechanistic evidence that can potentially influence the level of evidence reached, in particular in the absence of reliable randomised trial evidence. Finally, the process leading from evidence judgement to recommendations differs, though it is based on the same principles.

There are five possible levels of conclusion. In effect, the criteria define these terms, which appear in the matrices.

1. High certainty of an effect
2. Moderate certainty of an effect
3. Low certainty of an effect
4. Very low certainty of an effect
5. Substantial effect unlikely

An effect may be demonstrated directly by well designed, executed and analysed RCTs that address a relevant intervention and outcome; or inferred indirectly from RCTs with some

defects, or from relevant observational data, if supported by pertinent mechanistic data (see later section, page 15).
The following study designs are included in CUP Global Systematic Literature Reviews being conducted for studies of cancer survivors:

- RCTs with follow-up data of at least 6 months Follow-up of cancer cases from case-control studies
- Follow-up of cancer cases from cohort studies
- Cohort studies of cancer survivors
- Ancillary analyses from randomised controlled trials (RCTs)

High certainty (Strong evidence)

These criteria are for evidence strong enough to support a judgment of high certainty of a convincing effect.

1. Evidence of an effect from at least two well-designed independent RCTs with:
   - No substantial unexplained heterogeneity AND
   - No substantial evidence of publication bias

   (there should be at least two studies that show significant associations or a significant meta-analysis)

Moderate certainty (Strong evidence)

These criteria are for evidence strong enough to support a judgment of moderate certainty of an effect.

1. Evidence of an effect from at least two well-designed independent RCTs with:
   - Some unexplained heterogeneity allowed
   - No substantial evidence of publication bias

(there should be at least two studies that show significant associations or a significant meta-analysis)

OR

2. Evidence of an effect from any one of the following:
   - At least one well-designed RCT plus supportive evidence from cohort studies
   - Results from at least one well-designed pooling study of cohort studies
   - At least three well-designed cohort studies (ie potential to meta-analyse the data)

EACH WITH:
- No substantial unexplained heterogeneity
- No substantial publication bias
- Strong and plausible mechanistic evidence

Substantial effect unlikely

Evidence is strong enough to support a judgement that a particular intervention is unlikely to have a substantial effect on a cancer or other outcome.

Note: The Panel are asked to consider the time-frames being studied when drawing absence of effect conclusions.

All of the following are required:
Evidence of the absence of an effect from any of the following:

- at least two well-designed independent RCTs (there should be at least two studies that show significant associations or a significant meta-analysis)
- a well-designed pooling study of cohort studies that produces a summary estimate close to 1.0
- at least two well-designed cohort studies that produce estimates/summary estimates (for three or more studies meta-analysed) close to 1.0

Each with:

- No substantial unexplained heterogeneity
- Absence of a dose-response relationship (in follow-up studies)
- Absence of strong and plausible mechanistic evidence

Low certainty of an effect (weak evidence)

These criteria are for evidence that is too limited to permit a judgement of high or moderate certainty of an effect, but where there is suggestive evidence. The evidence may have methodological flaws, or be limited in amount, but shows a generally consistent effect. This level of evidence would not generally be used to justify making specific recommendations, though less rigid guidance might be developed.

1. Evidence from at least two well-designed independent RCTs, the confidence interval may include the null but associations are in a consistent direction with:

- Some unexplained heterogeneity allowed
- No substantial evidence of publication bias

OR

2. Evidence from one well-designed RCT but the confidence interval may include the null with:

- No substantial evidence of publication bias
- Strong and plausible mechanistic evidence

OR

3. Evidence of an association from a well-designed pooling study of cohort studies with:

- Some unexplained heterogeneity allowed

- No substantial evidence of publication bias

OR

4. Evidence from a well-designed pooling study of cohort studies, the confidence interval may include the null but associations are in a consistent direction with:
   - Some unexplained heterogeneity allowed
   - No substantial evidence of publication bias
   - Evidence for biological plausibility.

OR

5. Evidence of an effect from at least two independent cohort studies with:
   - No substantial unexplained heterogeneity
   - No substantial evidence of publication bias

   (there should be at least two studies that show significant associations or a significant meta-analysis)

Very low certainty of an effect (weak evidence)

Evidence is so limited that no firm conclusions can be made. Evidence may be judged 'very low certainty' for any of the following reasons:
- Too few studies available
- Substantial inconsistency in the direction of associations
- Poor quality of studies

Notes: There is currently no minimum length of follow-up for observational studies. It is important to note that people living with and beyond a cancer diagnosis can survive decades, and therefore the follow-up required may need to be longer for some cancers than others. *Published pooling studies* (ie those combining individual level data from different studies) are searched for by the team at Imperial College London and included in the Systematic Literature Reviews (SLRs) for consideration when evaluating the evidence.

Published meta-analyses (ie those statistically combining study level data) are searched for but not included in the SLRs. The reference lists are checked for completeness of the search and to provide information about the research landscape for the Panel.

All relevant data are synthesised in the literature reviews with data examined relative to time of diagnosis. Grading criteria are applied to the evidence for each exposure and outcome relationship in each timeframe of exposure assessment relative to time of diagnosis. Exposure assessment relative to period of treatment is highly desirable.

Upgrading and downgrading factors for incidence and survivors

Upgrading factors

These are factors that form part of the assessment of the evidence that, when present, can upgrade the judgement reached. An exposure that might be deemed 'limited – suggestive' causal factor in the absence, for example, of a biological gradient, might be upgraded to 'probable' if one were present. Alternatively, a 'low certainty' conclusion might be upgraded to 'moderate certainty'.

The application of these factors (listed below) requires judgement, and the way in which these judgements affect the final conclusion in the matrix are stated.

Factors in observational studies may include the following:

- Presence of a plausible biological gradient ('dose–response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be plausibly explained.
- A particularly large summary effect size (an odds ratio or relative risk of 2.0 or more, depending on the unit of exposure) after appropriate control for confounders.
- Consideration of precision in 1) measurement and 2) results from multiple studies. For 1), consideration of how accurately an exposure/outcome is measured within a study and for 2), consideration within and across studies of the level of statistical variability (for example, do studies exploring the same exposure-outcome pair produce similar results?).
- Consideration of precision within individual studies – a large effect size with poor precision (for example, a wide confidence interval), even if statistically significant, could reduce confidence in the finding.
- Evidence from randomised controlled trials in humans.
- Evidence from appropriately controlled experiments demonstrating one or more plausible and specific mechanisms operating in humans.
- Where mechanistic evidence in humans is limited/not available, robust and reproducible evidence from experimental studies in appropriate animal models, showing that relevant human exposures can lead to relevant cancer outcomes, can be considered.

The results of Mendelian randomisation studies are now included in reviews undertaken by Imperial College London and WCRF's other collaborators. When the results of these studies are supportive of the findings of epidemiological studies, they should be considered an upgrading factor. When they are unsupportive of such results, they should be considered a downgrading factor.

Downgrading factors

There are several factors to be taken into account when considering whether a study is well designed. Broadly, these relate to the design of the study in relation to the specific research question, execution of the study, and how the analysis was performed. The section below

details some of the issues, within these three areas, that should be considered when considering whether downgrading would be appropriate. These have been separated by study type, but many of them are relevant to all of the study types likely to be included in CUP Global reviews.

Features of well designed studies

The main features of well designed studies are described below. These are intended to provide guidance for the Panel. This is not intended to replace the Panel's expert judgement so a threshold is not provided to judge whether a study is 'well designed' or not.

CUP Global reviews include a risk of bias assessment, with the assessment tool depending on study type. These tools assess systematic error in studies and are used to provide information to support the Panel's evaluation of the design and reporting of studies.

Cohort studies

A 'well-designed' cohort study has a clearly defined scientific background and rationale for the investigation. The methods used within the study are appropriate, as well as clearly defined and reported. The key facets of the study include:

- Participant eligibility (including criteria/selection)
- Information about exposures/outcomes and other variables of interest (including confounders and how they will be handled). Within survivorship studies, examples of confounders include patient characteristics, tumour pathology and biological subtype, tumour stage, local and systemic treatments received. Stage will describe early (curable) vs advanced (non-curable) disease.
- Sources and assessment of data
- Sample size (including loss to follow-up)
- Potential sources of bias (risk of bias assessment is carried out using appropriate tools e.g. ROB-2)
- Statistical methods used (including subgroup, interaction or sensitivity analysis)

Pooling studies

Within this document 'pooling studies' relate to the pooling of individual, rather than study-level data (ie meta-analysis). Within the former, well-designed studies are those that use appropriate methods that are clearly defined and reported. This includes:

- Statistical methods used
- The populations under investigation
- How confounders and effect modifiers were handled
- Measurement error/validation
- Potential sources of error and bias
- How the individual-level data were standardised across studies

Randomised controlled trials (RCTs)

An RCT is generally considered well-designed when the following factors are clearly defined and reported:

- A clear a priori hypothesis – and how it is being tested
- How the study sample is appropriate for the hypothesis under investigation and how participants were chosen
- Inclusion/exclusion criteria
- How participants were randomised to the treatment group or intervention/control group (depending on the study design)
- Bias (selection and observer/information)
- Exposure and outcome-related information (measurement, precision etc)
- Confounding factors
- How the treatment of the participants may have differed between the intervention and control group
- The type of blinding that occurred (e.g., double blinding)
- Intention to treat analysis – that those assigned to a treatment group are analysed within the group, even if they did not receive the treatment
- Drop-out rate
- Sample size and power calculation

Two further factors to consider when assessing RCTs are:

1) The generalisability of the RCT participants (e.g., in terms of demographic and socioeconomic factors) to the general population of interest
2) The analysis undertaken and reported is focused upon the initial research question/hypothesis rather than from secondary or subgroup analysis.

Note: Occasionally for specific cancers and populations, where there is limited data from cohort studies, we may need to use case control studies. These have limitations which will need to be considered when reviewing the evidence.

Special considerations regarding cancer survivor studies

1. A greater weight is placed on RCTs versus follow-up studies for the grading criteria for cancer survivors compared with the grading criteria for cancer incidence. This is because of the greater possibility and difficulty in correcting for, confounding in observational studies. Evidence of an effect from a meta-analysis of RCTs, or at least two well-designed independent RCTs, is required for evidence to be judged 'convincing'.
2. RCTs can also determine adverse effects. Most treatment trials include careful attention to adverse effects, and that needs to be addressed for nutrition/physical activity/weight change trials also.
3. When good quality data from RCTs are available, strong and plausible mechanistic evidence is desirable, but is not required, for evidence to be judged 'convincing'.
4. RCT evidence is not required for evidence to be judged 'probable' - but strong and plausible mechanistic evidence is required if there is not good RCT evidence. The observational data needs to be fully adjusted for potential confounders, such as the tumour type, type of treatment, amount of treatment received, and the dissemination of the disease.
5. RCT evidence may have good internal validity if it is well conducted; however, patients included in RCTs may not be representative of the wider population of

cancer survivors. Survivors who do not enter RCTs may be sicker and/or have different underlying factors and risks and could, therefore, have lower survival. Cohort studies with large case numbers and a high response rate may have greater generalisability than RCTs, but the selection of participants into cohort studies of cancer studies will also influence study generalisability so must be considered.

6. Survivorship studies are often embedded in trials, case-control or cohort studies so may be based on secondary analyses from these studies. In this case, there may not be an apriori hypothesis.

Special considerations to take into account when grading cancer survivor evidence

These were developed for use with the breast cancer survivors reviews and have not yet been used for reviewing other cancer types. These will be kept under review.

1. ***What treatments have the cohort members had?*** Treatment varies by tumour type and patient characteristics. The type and amount of treatment can have greater effect on survival than most exposures related to diet, nutrition, body weight, and physical activity, and there is likely confounding factors. In the United States, for example, access to treatments varies by sociodemographic status (location, socioeconomic status, race and ethnicity), as does diet and physical activity, so an apparent diet-survival relationship may be confounded by the type of treatment received. This also pertains to stage at diagnosis, but stage is more easily ascertained in studies and is thus easier to control for than treatment information.
2. ***Healthy cohort effect.*** Some types of cancer recur early and cause early mortality. If a survivor cohort is assembled a long time after diagnosis, individuals at high risk for mortality may not be included. This has happened in some cohorts already (including the HEAL study), and in any trial that included persons diagnosed in the more distant past (for example the WHEL study). This is particularly important for some types of cancer (such as breast cancer negative for oestrogen and progesterone receptors and HER2).
3. ***Time periods and changes in treatments.*** Due to improved knowledge regarding tumour type, new treatment regimens have changed the expected effect of treatment and thus cancer mortality (for example, 15-20% of breast cancer cases are now known to be positive for HER2). Treatment regimens vary according to time periods, country, and socio-economic status within countries.
4. ***Early mortality vs. late mortality.*** For most cancer types, independent of tumour type, early recurrence is that occurring within the first 2 years (possible due to already metastatic disease not responding to neo/adjuvant treatment). For the breast cancer survivor reviews, 10-year and, to a lesser extent, 5-year breast cancer survival was specifically discussed. This may be revised for other cancer types. This underlines the importance of understanding breast cancer as a chronic disease with longer expected survival time.

Mechanistic evidence in CUP Global reviews

Cancer incidence and biological mechanisms

A team of researchers at IARC are undertaking reviews of the biological mechanisms that underpin the results found within the CUP Global epidemiological reviews. This work builds on the evidence that was provided in this regard to support the Third Expert Report. The team, alongside WCRF, have been working to produce a protocol which operationalises the

review of this type of evidence. Within the protocol the following definition of mechanisms is included:

> *Mechanisms evaluated using this framework are defined as biological processes linking diet, nutrition, physical activity and cancer hallmarks/incidence and cancer-specific survival. Other mechanisms, including comorbidities, treatment factors (dose, completion etc) will not be included within the mechanistic reviews but may be part of the epidemiological reviews undertaken by Imperial College London and others.*

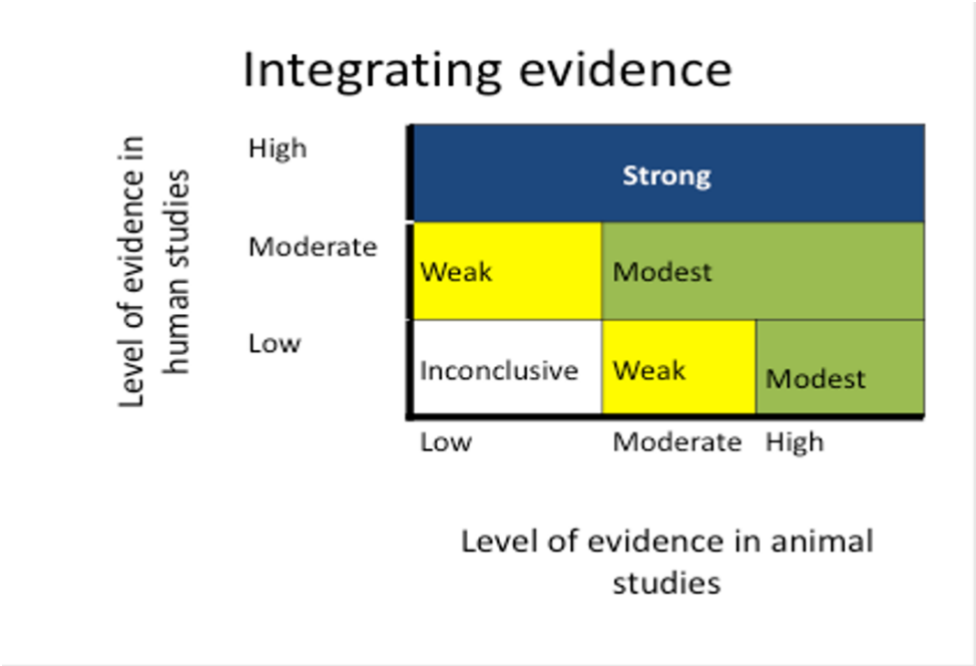
At present, CUP Global does not collect or report information related to the wider structural external factors, such as the social determinants of health, that impact cancer outcomes. WCRF is beginning to consider how best to start doing so, as well as information relating to health inequalities. To progress this work, and develop our understanding of how it relates to both the epidemiological reviews and the biological mechanistic evidence, the opinion of the expert committees and Panel will be sought.

Within CUP Global, mechanisms-related systematic reviews are undertaken to 1) link an exposure with intermediate phenotypes in the mechanistic pathway of interest and 2) link the intermediate phenotypes with the cancer outcome under investigation. The mechanistic reviews currently utilise both a semi-systematic and narrative approach, depending on the needs of the investigation. Given the innovative nature of the work, the methodology is constantly reviewed and amended to make sure that it is both systematic and pragmatic in its approach to synthesising complex mechanistic evidence. The results of the mechanisms reviews are then included in the grading criteria, as well as considered an upgrading factor.

This will include assessment of:

- Study design
- Risk of bias (selection and publication)
- Relevance
- Precision
- Consistency in results
- Magnitude of effect
- Whether a dose-response was found
- Confounders

While priority is given to human studies, animal models will also be included (where appropriate and possible) and the strength of the overall body evidence will be considered [4].



Cancer survivorship and mechanisms

Similarly to the work undertaken to support the epidemiological reviews related to cancer incidence, the IARC team will review the biological mechanisms underpinning the associations found within CUP Global cancer survivor reviews.

To support the survivors work, WCRF will also need to develop an understanding of the wider clinical and socioeconomic factors that influence cancer-related, as well as other, outcomes among those living with and beyond cancer. This includes the presence of comorbidities, cancer stage, social determinants of health, health inequalities, and treatment exposures (e.g., treatment dose and completion). This work is beyond the scope of the reviews undertaken by IARC, and a different approach will be developed.

Context for using the criteria

The ultimate aim of the criteria is to provide a standardised basis for the Panel's recommendations and guidance.

For cancer survivors, high certainty of an effect of a specific intervention can only be reached with well designed, executed and analysed RCTs. Lower levels of certainty may be supported by RCTs with acknowledged limitations, or with observational data, but only if supported with mechanistic evidence and where the impact of potential biases have been carefully considered.

For cancer incidence, there are other considerations. The overall aim is to provide recommendations on ways of living that reduce cancer risk. However, compared with RCTs among cancer survivors, RCTs may be neither ethical, practical nor definitive for cancer prevention. The need for such criteria, therefore, rests on two issues: first, the relevance of particular study designs to illuminate the questions of importance, and second, the impact

on cancer risk of exposures that are complex, that are difficult to manipulate, and that have their effect over decades or whole life spans, or even across generations. This is analogous to the situation described by Bradford Hill when addressing occupational exposures that would not be susceptible to testing through the most robust form of evidence, RCTs. When faced with a problem of public health significance, but without the possibility to test the causality of observed associations by RCTs, Bradford Hill noted that a best judgement needed to be made to allow consideration of public health interventions.

Bradford Hill was aware not only of the power of RCTs to vigorously test the effects of interventions or the causality of associations, but also of limitations in their application in certain contexts. Although the results of well-conducted and well-executed RCTs provide robust answers to certain questions, not all hypotheses can be tested. Many questions of clinical importance can be directly tested, but controlled manipulation of diet and physical activity over a lifetime is clearly not readily amenable to such testing. RCTs have good internal validity (they give a robust answer to the question tested) but may have poor external validity (the question able to be tested is not directly applicable to the real-life situation). This may be for a range of reasons, for instance because of the use of atypical populations (such as those selected for high risk), or abnormal exposures (such as high- dose supplements), or simply because over the long term of an extended RCT differential attrition and adherence between the test and control groups mitigates the ability to ascribe differences in outcome to the test intervention. Although RCTs still inform overall judgements, the directness of the relationship of their results to the question of relevance may vary.

In contrast, prospective observational studies offer an opportunity to identify characteristics in real populations that are associated with real outcomes over long periods of time, although few cohort studies have lifetime exposure information. There is though, the potential for confounding, which means that ascribing causality may be problematic. In practice, when faced with an important public health issue such as the impact of diet, body weight and physical activity on cancer, where public guidance is needed, that guidance needs to be based on the best evidence, and this comes from a variety of sources (including observational data, RCTs and experimental evidence in laboratory models). When evaluating mechanistic evidence, WCRF will apply a hierarchy in terms of strength of evidence, with human studies considered to provide stronger evidence than animal studies. The criteria offer a framework within which the process of assessing the evidence, drawing conclusions and making recommendations can take place.

Inevitably, it is problematic to attempt to isolate the impact of individual components of complex patterns of exposure, within which many components are associated with each other (for example, because they occur in the same foods). Therefore, there is an even greater degree of certainty that the truly causal factor or factors lie within the identified broad pattern of exposure (that is, all the conclusions together), than there is for each singular component.

We also need to consider the issue of competing risks. For example, 1) an exposure that causes a particular disease at an earlier age might lead to fatalities or 2) treatments that

impact the ability to see risk for another disease (e.g., specific type of cancer) – risk might be unmasked if treatments change risk of death from the earlier occurring disease.

Although RCTs are now regarded as the norm for supporting clinical interventions, it is important to recognise that even in this setting, extrapolations from the evidence are usual in the case of individual patients (even if this is not always explicitly recognised by practitioners). For instance, the rigorous selection criteria in high-quality RCTs means that typical patients, who often have more than one condition, may not be eligible to participate. This rightly does not stop the practitioner applying professional judgement in a particular case, in the face of a degree of uncertainty about the evidence. Therefore, both clinical and public health practice rely on the recognition of the need for an intervention or guidance (for example, because of symptoms in patients or because of a preventable public health problem) and the application of professional judgement to a particular case in the face of incomplete evidence.

Proof is a mathematical construct, and certainty is rarely attainable in biology. Therefore, when dealing with degrees of uncertainty it is important to be methodical about specifying the level of confidence expected to support an intervention or recommendation/guidance. That is what these criteria do. It is also important to consider that RCTs are powered to see effects that are clinically significant, but small effects might have important public health impact (i.e., effect cancer incidence rates in a population) if the exposure is common (and therefore has an appreciable attributable fraction).

Food-based approach

Terms used in the text of the Third Expert Report reflect the Panel's decision that its judgements and recommendations should, whenever possible, be based on foods and drinks rather than on nutrients or other bioactive constituents. This is in part because dietary constituents associated with foods are grouped with these foods. Thus, matrix entries in Exposures: Wholegrains, vegetables and fruit list 'foods containing dietary fibre' (rather than dietary fibre), and in Exposures: Alcoholic drinks list 'alcoholic drinks' (rather than alcohol or alcohol as ethanol).

The food-based approach is also justified because of the uncertainty that any food constituent is a true causal factor, rather than simply a marker for the particular foods in which it is found or for other dietary constituents found in the same foods, or for other associated health-related factors. In Exposures: Other dietary exposures, some supplements of micronutrients appear in matrices graded as 'convincing' or 'probable'. These judgements are derived from the findings of good-quality RCTs, sometimes also supported by observational studies, clearly showing that supplements of these micronutrients – rather than the foods containing them – affect the risk of cancer; as, for example, with beta-carotene supplements and increased risk of lung cancer.

Sometimes the studies that are the basis for the Panel's work have used markers of exposure. Many epidemiological studies use BMI, waist circumference and waist- hip ratio as markers of body fatness. When there is clear evidence of an underlying mechanism for

body fatness, the Panel has agreed that the term ‘body fatness’ best represents the causal factor (see Exposures: Body fatness and weight gain).

As exceptions to this approach, the Panel has made judgements on ‘adult attained height’ and ‘greater birthweight’, as shown in the matrices. Many epidemiological studies have reported on height and birthweight. It is thought that associations between height and cancer risk reflect some causal association with a combination of genetic, environmental, hormonal and nutritional growth factors affecting growth during the period from preconception to completion of linear growth. Uncertainty as to the precise mechanisms underlying the observations with ‘adult attained height’ and ‘birthweight’ mean that the Panel was not able to determine the appropriate causal factors to be shown in the matrices (see Exposures: Height and birthweight). Instead, the anthropometric markers have been included, with appropriate footnotes.

CUP Global matrices

The matrices display the Panel’s judgements on whether particular aspects of diet, nutrition and physical activity may modify (or not modify) the risk of cancers of specific sites. Necessary clarifications and qualifications are stated in footnotes to the matrices. In some cases, analysis may show that any association or effect begins or ends, or is less apparent, below or above evident ‘thresholds’. For example, alcoholic drinks appear to increase the risk of some cancers (such as liver and colorectal) only above certain levels of consumption. Such amounts are specified in a footnote to the relevant matrices. When matrices include no such footnotes (as for alcohol and postmenopausal breast cancer), this is because no lower or upper threshold for the association or effect has been identified. In such cases, matrix entries showing or suggesting a causal association should be taken to mean that the association or effect is across the whole range of dietary intake, amounts of physical activity or degrees of body fatness found in the studies analysed.

Figure 1: Example of a summary matrix: physical activity and the risk of cancer

2018		DECREASES RISK		INCREASES RISK	
		Exposure	Cancer site	Exposure	Cancer site
STRONG EVIDENCE	Convincing	Physical activity[1]	Colorectum (colon) 2017[2]		
	Probable	Physical activity[1]	Breast (postmenopause) 2017[3] Endometrium 2013[5]		
		Vigorous intensity physical activity	Breast (premenopause) 2017[3] Breast (postmenopause) 2017[3]		
LIMITED EVIDENCE	Limited - suggestive	Physical activity[1]	Oesophagus 2016[4] Lung 2017 Liver 2015 Breast (premenopause) 2017[3]	Sedentary behaviours	Endometrium 2013[5]
STRONG EVIDENCE	Substantial effect on risk unlikely	None identified			

1 The exposure of physical activity includes evidence for all types of activity and all intensity levels.
2 The evidence for physical activity and colorectum is for colon cancer only – no conclusion was drawn for rectal cancer.
3 In addition to physical activity, there was sufficient evidence for the Panel to make a separate judgement for vigorous intensity physical activity and breast cancer (pre and postmenopause).
4 The evidence for physical activity and oesophageal cancer includes unspecified, adenocarcinoma and squamous cell carcinoma.
5 The evidence for sedentary behaviours and endometrial cancer was marked by sitting time.

References

[1] Giovannucci, E. Nutritional epidemiology and cancer: A Tale of Two Cities. *Cancer Causes Control* **29**, 1007–1014 (2018). https://doi.org/10.1007/s10552-018-1088-y

[2] Hill AB. The environment and disease: association or causation? Proceedings Royal Society Medicine. 1965; 58: 295-300

[3] World Cancer Research Fund/ American Institute for Cancer Research. Food, Nutrition, Physical Activity, and the Prevention of Cancer: a Global Perspective. 2007

[4] Sarah J. Lewis, Mike Gardner, Julian Higgins, Jeff M.P. Holly, Tom R. Gaunt, Claire M. Perks, Suzanne D. Turner, Sabina Rinaldi, Steve Thomas, Sean Harrison, Rosie J. Lennon, Vanessa Tan, Cath Borwick, Pauline Emmett, Mona Jeffreys, Kate Northstone, Giota Mitrou, Martin Wiseman, Rachel Thompson, Richard M. Martin; Developing the WCRF International/ University of Bristol Methodology for Identifying and Carrying Out Systematic Reviews of Mechanisms of Exposure–Cancer Associations. *Cancer Epidemiol Biomarkers Prev* 1 November 2017; 26 (11): 1667–1675. https://doi.org/10.1158/1055-9965.EPI-17-0232

Appendices

Appendix 1- Summary of the grading criteria for cancer incidence

Evidence grades		GRADING CRITERIA FOR EVIDENCE ON DIET, NUTRITION, PHYSICAL ACTIVITY AND CANCER INCIDENCE	Het	PB	Mec
Strong evidence	**Convincing**	Evidence from more than one good quality study type, including at least two independent cohort studies (acceptable study designs listed below)*	No	No	Required
	Probable	Evidence from at least two good quality independent cohort studies	No	No	Required
Limited evidence	**Limited suggestive**	Evidence from at least two independent cohort studies	Yes	Yes	Required
	Limited – no conclusion	Any of the following reasons: - Too few studies available - Inconsistency of direction of effect - Magnitude of effect unlikely to affect cancer risk - Poor quality of studies (for example, lack of adjustment for known confounders)	-	-	-
Strong evidence	**Substantial effect on risk unlikely**	Evidence of the absence of an effect. All of the following generally required: - Evidence from more than one good quality study type (acceptable study designs listed below)* - Evidence from at least two independent cohort studies - Summary estimate of effect close to 1.0 for comparison of high- versus low-exposure categories	No	No	Absence

Het: Substantial unexplained heterogeneity or some unexplained heterogeneity
PB: Publication bias
Mec: Strong and plausible mechanistic evidence is required, desirable but not required, not required, or absent
*RCTs, longitudinal, observational, or pooled analyses of individual data of these studies. Good-quality studies exclude with confidence the possibility that the observed association results from random or systematic error, including *confounding*, measurement error and *selection bias*.

Special upgrading factors:
- Presence of a plausible biological gradient ('dose response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly. This is also a requirement for the convincing (strong evidence) grade.
- A particularly large summary effect size (an odds ratio or relative risk of 2.0 or more, depending on the unit of exposure), after appropriate control for confounders.
- Consideration of precision.
- Evidence from randomised trials in humans.
- Evidence from appropriately controlled experiments demonstrating one or more plausible and specific mechanisms operating in humans.
- Robust and reproducible evidence from experimental studies in appropriate animal models showing that typical human exposures can lead to relevant cancer outcomes.

Appendix 2- Summary of the grading criteria for cancer survivors

Evidence grades		GRADING CRITERIA FOR EVIDENCE ON DIET, NUTRITION, PHYSICAL ACTIVITY AND CANCER SURVIVAL	Het	PB	Mec
Strong evidence	**High certainty of an effect**	Evidence of an effect from at least two well-designed independent RCTs	No	No	Not required
	Moderate certainty of an effect	Evidence from at least two well-designed independent RCTs	Some	No	Not required
		OR Evidence from one well-designed RCT plus evidence from well-designed cohort studies	No	No	Required
		OR Evidence from at least one well-designed pooling study of cohort studies	No	No	Required
		OR Evidence from at least three independent well-designed cohort studies (i.e. potential to meta-analyse the data)	No	No	Required
Weakevidence	**Low certainty of an effect**	Evidence from at least two well-designed independent RCTs,the confidence interval may include the null but associations are in a consistent direction	Some	No	Not required
		OR Evidence from one well-designed RCT but the confidence interval may include the null	No	No	Required
		OR Evidence from a well-designed pooling study of cohort studies	Some	No	Not required
		OR Evidence from a well designed pooling study of cohort studies, the confidence interval may include the null but associations are in a consistent direction	Some	No	Required
		OR Evidence from at least two independent cohort studies	No	No	Not required
	Very low certainty of an effect	Any of the following reasons: - Too few studies available - Substantial inconsistency in the direction of associations - Poor quality of studies	-	-	-
Strong evidence	**Substantial effect unlikely**	Evidence of the absence of an effect from any of the following: a) At least two well-designed independent RCTs (estimates close to 1) b) A well-designed pooling study of cohort studies that produces a summary estimate close to 1.0 d) At least two well-designed cohort studies that produce estimates/summary estimates (3 or more studies) close to 1 - Absence of a dose response relationship (in follow-up studies) Note: the Panel are asked to consider the time-frames behind studied when drawing absence of effect conclusions.	No	-	Absence

Het: Substantial unexplained heterogeneity or some unexplained heterogeneity
PB: Publication bias
Mec: Strong and plausible mechanistic evidence is required, desirable but not required, not required, or absent
Special upgrading factors:
- Presence of a plausible biological gradient ('dose response') in the association. Such a gradient need not be linear or even in the same direction across the different levels of exposure, so long as this can be explained plausibly.
- A particularly large summary effect size (a relative risk of 2.0 or more, or 0.5 or less, depending on the unit of exposure), after appropriate control for confounders.
- Evidence from appropriately controlled experiments demonstrating one or more plausible and specific mechanisms.
- All plausible known residual confounders or biases including reverse causation would reduce a demonstrated effect, or suggest a spurious effect when results show no effect. Special considerations important for evidence for breast cancer survivors including the following potential confounding variables – the type of tumour, type of treatment, amount of treatment received, and the dissemination of the disease.

Appendix 3 - Process for developing Recommendations and Guidance for cancer survivors