Questions for Evidence Extraction

Meta-analysis answers a research question of the form: "What is the impact of [some predictor or intervention of interest] on [some outcome of interest]?" Your research question is,
"What is the impact of a vegetarian diet on the overall mortality rate due to all causes?"

This form guides the process of extracting evidence from written reports of research findings for meta-analysis. It has two parts. The first part documents study-level information which includes details about each written report.

Study-Level Information

In this section, you will document information about the written report you are coding. ***Please feel free to use shorthand or copy and paste blocks of text from the article itself.***

Study Identity

1. What is the title of the report?

2. List the authors of the study.

3. What year was the study published?

General notes:

Study Context

Please use the following items to document the study design and the context of data collection, paying special attention to information that might impact how you interpret the study results.

1. What kind of study is this?
    - ❏ Experimental (i.e., participants randomly assigned to groups)
    - ❏ Observational (i.e., participants' behavior was observed without intervention) ↓

      Did the study use a control group to establish a baseline measurement of the outcome without intervention?
        - ❏ Yes
        - ❏ No ↓

          You've indicated that this study did not compare the intervention of interest to a control condition. This means you will not be able to include this study in a meta-analysis because the effect sizes that are combined in meta-analysis are relative measurements of the impact of an intervention versus baseline. It may still be useful for you to consider this study in your review. However, note that the lack of a control condition often signals that a study is poorly designed.

      ~~The groups of interest for the study are defined based on:~~
        - ~~❏ Exposure to the intervention of interest (i.e., cohort study)~~
        - ~~❏ Exposure to the outcome of interest (i.e., case-control study)~~
    - ❏ Other kind of study ↓

      Name: _____

      Description: _____

2. What was the study setting?
    - ❏ Controlled (e.g., clinic, laboratory)
    - ❏ Naturalistic (e.g., home, school, work)
    - ❏ Crowdsourced (e.g., webpage)
    - ❏ Other setting ↓

      Name: _____

      Description: _____

General notes:

Participants

Please use the following items to describe participants in the study and how they were selected.

1. Briefly describe the process used to recruit participants for the study.

2. List any inclusion or exclusion criteria that were used to filter out participants in the study.

3. Briefly describe the **final sample of participants** after the recruitment and exclusion processes in terms of relevant characteristics (e.g., demographics, occupation, health status). These are the people whose data are reflected in the results of the study.

Result-Level Information

In this section, you will document information about each estimate of the effect of interest in the written report you are coding. Please fill out the questions in this section for each reported result you would like to include in your meta-analysis.

Note: Be careful not to create duplicate entries if the same results are reported in separate written reports.

How many results from this written report would you like to extract for meta-analysis?

Measurement

Please use the following items to describe how the outcome of interest was defined and measured in the study (i.e., how the outcome of interest is operationalized).

1. What was the independent variable in the study? We need a shorthand for the intervention of interest that was compared to a control condition or baseline.

2. How was the independent variable defined? Briefly describe how the author(s) implement the intervention and control conditions in the study.

3. What was the dependent variable in the study? We need a shorthand for the outcome of interest that was measured.

4. How was the dependent variable defined? Briefly describe how the author(s) measure or observe the outcome of interest in the study.

General notes:

5. Was the outcome of interest dichotomous (i.e., an event that did or did not happen) or continuous (i.e., a scale with more than two possible values)?
    - ❏ Dichotomous.
    - ❏ Continuous.
6. Was the outcome measurement taken **between-subjects** such that there is only one observation per participant, or was the outcome measurement repeated **within-subjects** over time such that there are multiple observations for each individual participant?
    - ❏ The measurement was taken once per participant **between-subjects**.
    - ❏ The measurement was repeated **within-subjects**.
7. What was the time schedule for measurement in the study? When were measures taken?
    - ❏ The measurement was taken at **one point in time**. ↓

      Do the authors report the point in time when the measurement was taken (e.g., duration after the intervention of interest)?
        - ❏ Not applicable.
        - ❏ No.
        - ❏ Yes. ↓

          Record the time: _____

          What unit of time is reported? _____
    - ❏ The measurement was taken at **multiple points in time**. ↓

      Do the authors aggregate measurements over time and present one overall measure, or do the authors report the measure at multiple durations after the intervention of interest?
        - ❏ The measurement is **aggregated over time**.
        - ❏ The measurement is reported for **multiple points in time**. ↓

          The authors may have reported on more points in time than you want to include in your meta-analysis. Use your discretion. Which points in time or durations after the intervention are you interested in documenting and potentially including in your meta-analysis?

          List the times: _____

          What unit of time is reported? _____

General notes:

8. Do the authors report this measure separately for different groups of participants (e.g., grouping their sample by demographics or some other factor)?
    - ❏ Group effects were **not reported**.
    - ❏ Group effects were **reported**. ↓

      The authors may have reported on more group comparisons than you want to include in your meta-analysis. Use your discretion. Which group-level effects are you interested in documenting and potentially including in your meta-analysis?
        - ❏ I am only interested in the effect **aggregated across groups**.
        - ❏ I am interested in the effect in **different groups**. ↓

          List the grouping factors that are of interest (e.g., gender): _____

          For each factor listed above, what levels of the factor were reported? _____

General notes:

Effect Size

Please use the following items to capture information reported about estimates of effect size.

You will almost certainly find this information in the Results section.

1. Do the authors report sample size separately for the intervention and control groups, or do they only report overall sample size for the whole study?
    - ❏ Sample size reported separately for intervention and control groups.
    - ❏ Only overall sample size reported (assume a balanced design).
    - ❏ Neither. ↓

      Without knowing sample size, we cannot calculate effect size.
2. Do the authors report the count or proportion of individuals who experience the outcome of interest in the intervention and control groups?
    - ❏ Count.
    - ❏ Proportion.
    - ❏ Neither. ↓

      Without knowing the count or proportion of participants in the intervention and control groups who experience the outcome of interest, we cannot calculate effect size for a dichotomous outcome.
3. Do the authors report the mean outcome separately for the intervention and control groups, or do they report the mean difference between the intervention and control groups?
    - ❏ **Mean outcome** for intervention and control groups.
    - ❏ **Mean difference** between intervention and control groups.
    - ❏ Neither. ↓

      Without knowing the mean difference between intervention and control groups, we cannot calculate effect size for a continuous outcome.

General notes:

4. How do the authors report the reliability of the estimated difference between the intervention and control groups? Please check one of the following.
    - ❏ Standard deviations reported separately for intervention and control groups.
    - ❏ Pooled standard deviation.
    - ❏ Standard error.
    - ❏ t-value.
    - ❏ F-value.
    - ❏ p-value. ↓

      Was this a 1- or 2-tailed test?
        - ❏ 1-tailed test.
        - ❏ 2-tailed test.
    - ❏ Confidence Interval (CI). ↓

      What is the confidence level? _____%
    - ❏ None of the above ↓

      Without one of the above measures of reliability, we cannot calculate effect size for a continuous outcome.

5. If a statistical model produced the results, what covariates (if any) were adjusted for? These are factors which the authors controlled for when calculating effect size.

General notes:

6. Use this table to document the numerical effect size information reported in the study. Note that the authors may report more comparisons than you will include in your meta-analysis. You will only be able to aggregate effect sizes from these comparisons if the same comparison is made in multiple studies. **To save time, you should only fill in numbers which seem necessary to answer your research question**.

Possible columns:

- 'Condition'
    - Show if answer to Measurement 5 is dichotomous or if user gives first answer under Effect Size 1, 3, or 4
    - Auto-populate two rows: Intervention vs Control
- 'Sample Size (n)'
    - Always show
- 'Count' or 'Proportion'
    - Show if answer to Measurement 5 is dichotomous AND depending on answer to Effect Size 2
    - Show exactly one of these two columns
- 'Mean' or 'Mean Difference'
    - Show if answer to Measurement 5 is continuous AND depending on answer to Effect Size 3
    - Show exactly one of these two columns
- 'Standard Deviation', 'Pooled Standard Deviation', 'Standard Error', 't-value', 'F-value', 'p-value (__-tails)', or '___% Confidence Interval'
    - Show if answer to Measurement 5 is continuous AND depending on answer to Effect Size 4
    - Show exactly one of these possible columns for reliability information
- Group
    - Show if the user answers that they are interested in different groups in Measurement 8.
    - Auto-populate rows from levels of factors elicited from the user.
- Time Point
    - Show if the user answers that they are interested in multiple points in time in Measurement 7.
    - Auto-populate rows from time points elicited from the user,

---

General notes:

Evidence Quality Assessment

Risk of Bias

Please use these questions to gage how well the study was conducted (i.e., internal validity). The subsections below reflect various threats to validity which could undermine the fidelity of effect size estimates.

Selection Bias

1. Did the recruitment process result in intervention and control groups that were systematically different?
    - ❏ Yes
    - ❏ No
    - ❏ Not sure

    This could be an issue if, for example:

    - *Intervention and control groups were recruited using different procedures or inclusion criteria.*
    - *Participants in the study were volunteers.*
    - *The non-response rate to the recruitment message was different between the intervention and control groups.*
    - *The intervention and control groups were not recruited from the same population.*
    - *The intervention and control groups were recruited at different times.*
2. Did exclusion or attrition result in intervention and control groups that were systematically different?
    - ❏ Yes
    - ❏ No
    - ❏ Not sure

    This could be an issue if, for example:

    - *Participants who withdrew from the study were systematically different from the final sample of participants.*
    - *Participants who were excluded from the study were systematically different from the final sample of participants.*
    - *The study authors handled missing data in a way that relied on strong and potentially unreasonable assumptions (e.g., an unreasonable imputation strategy).*
3. Did the study fail to randomly assign participants to intervention and control conditions?
    - ❏ Yes
    - ❏ No
    - ❏ Not sure

    This could be an issue if, for example:

- *The authors did not use a random or pseudo-random process for group assignment.*
- *The assignment process was predictable to participants or study personnel.*
- *The study was not an experiment.*

4. Did the study fail to control for confounding variables that could explain differences in the outcome of interest between the intervention and control groups?
   - ❏ Yes
   - ❏ No
   - ❏ Not sure

   This could be an issue if, for example:
   - *The authors failed to measure known confounding variables.*
   - *Potentially confounding variables were differently distributed within the intervention and control groups.*
   - *The study authors did not control for known confounding variables by either matching the study groups on them or using them as covariates in a regression.*
   - *The study authors attempted to control for a variable that was **not** a confounding factor in the intervention-outcome relationship (e.g., by overmatching or including a collider in their regression model).*

Measurement Issues

1. Did the study fail to isolate the impact of being in the intervention versus control group?
   - ❏ Yes
   - ❏ No
   - ❏ Not sure

   This could be an issue if, for example:
   - *Participants didn't stay in the group they were assigned to (e.g., by not adhering to a treatment schedule).*
   - *The procedure for determining membership in the intervention and control groups involved a subjective judgment on the part of study personnel.*
   - *The procedure for determining membership in the intervention and control groups involved unverified recall and/or self-report on the part of the participant.*
   - *Study personnel were not blind to group membership and treated participants differently depending on whether they were in the intervention or control condition.*
   - *Participants were not blind to group membership and changed their behavior to satisfy the expectations of study personnel.*

2. Did the study fail to detect or measure the outcome of interest?
    - ❏ Yes
    - ❏ No
    - ❏ Not sure

    This could be an issue if, for example:

    - *Different procedures were used for measuring the outcome of interest in the intervention and control groups, respectively.*
    - *The procedure for measuring the outcome of interest involved a subjective judgment on the part of study personnel.*
    - *The procedure for measuring the outcome of interest involved unverified recall and/or self-report on the part of the participant.*
    - *The outcome of interest was already present at the start of the study (e.g., participants in a study on weight loss were already at a healthy weight).*
    - *The follow-up period in which measurements were made was not long enough to allow the outcome of interest to occur (e.g., a study of cancer risk that only measures the presence of cancer for two months after intervention).*
    - *Participants were previously exposed to conditions that would impact the outcome of interest.*

Reporting Bias

1. Did the study authors seem to selectively report findings to support a particular narrative?
    - ❏ Yes
    - ❏ No
    - ❏ Not sure

    This could be an issue if, for example:

    - *Not all outcomes described in the study's methods were reported in the results.*
    - *The study was funded or run by an entity that has an incentive to report a particular result.*
    - *The authors did not adequately describe their procedures for data collection and analysis such that an independent party to reproduce them.*
    - *The study authors seem to have stopped data collection early for benefit (e.g., once they obtained a statistically significant result).*

Applicability

Please use these questions to gage how well the study applies to the situation about which you are trying to make an inference (i.e., external validity). The subjections below reflect various concerns which could limit the generalizability of the study findings.

1. Was the study conducted under conditions that differ in important ways from the situation you are trying to make an inference about?
    - ❑ Yes
    - ❑ No
    - ❑ Not sure
2. Did the sample of participants differ in important ways from the population you are trying to make an inference about?
    - ❑ Yes
    - ❑ No
    - ❑ Not sure
3. Was the intervention evaluated in this study different in important ways from the intervention you are interested in making an inference about?
    - ❑ Yes
    - ❑ No
    - ❑ Not sure
4. Did the study compare the intervention of interest to a control condition that is different in important ways from what would happen if you didn't intervene in the situation you are trying to make an inference about?
    - ❑ Yes
    - ❑ No
    - ❑ Not sure
5. Was the outcome measured in this study different in important ways from the outcome you are interested in making an inference about?
    - ❑ Yes
    - ❑ No
    - ❑ Not sure

    This might be an issue if, for example:
    - *The follow-up duration in which outcomes were measured was not similar to the timeframe about which you are interested in making an inference.*
    - *The study measured an outcome which was not of practical importance but was thought to be a proxy for the outcome of interest.*

Coding Manual

This manual describes the evidence we are looking for in each question of the evidence extraction form. Reference these descriptions for clarification and consistency as needed.

Study Context

1. What kind of study is this?

   Description: We differentiate kinds of studies depending on the degree of control the authors exert on the situation under study. In experiments, the authors may assign participants to treatment and control groups and attempt to isolate the effect they are measuring through experimental control. In observational studies, the authors measure the situation under study in the real-world without exerting their influence. In most observational studies, there should still be a control group whose experience reflects the absence of the intervention or treatment of interest without random assignment.
   Location: Articles will often explicitly state what kind of study was run, usually in the Method section.
   Importance: When comparing studies, we will want to consider whether different kinds of studies yield different patterns of results.

2. What was the study setting?

   Description: Studies present data gathered in different kinds of settings or environments. A controlled setting is one in which the authors have attempted to isolate the effect of interest by eliminating factors that might exist in real-world situations. In contrast, a naturalistic setting is one in which the authors have exerted no influence and are taking measurements in the real world. A crowdsourced setting is one in which data is collected online.
   Location: Study settings will often align with the kind of study and should be reported in the Method section.
   Importance: When comparing studies, we will want to consider whether different settings yield different patterns of results.

Participants

1. Briefly describe the process used to recruit participants for the study.

   Description: We think of participants in studies as being sampled from a particular population. Populations are essentially groups defined by common characteristics. If the recruitment process in a study biases the sample to include or omit a particular kind of participant (i.e., selection bias), we want to document this information so we can reason about the population in the study.
   Location: Information about the recruitment process will often be reported in the Method section.

Importance: The effect we are interested in might be different in different populations, so we will want to consider whether the studies we plan to combine are sampling such different populations that it doesn't make sense to aggregate their results in a meta-analysis.

2. List any inclusion or exclusion criteria that were used to filter out participants in the study.
   Description: We think of participants in studies as being sampled from a particular population. Populations are essentially groups defined by common characteristics. If a study excludes participants in a way that biases the sample to omit a particular kind of participant (i.e., selection bias), we want to document this information so we can reason about the population in the study.
   Location: Information about exclusions will often be reported in the Method or Results section.
   Importance: The effect we are interested in might be different in different populations, so we will want to consider whether the studies we plan to combine are sampling such different populations that it doesn't make sense to aggregate their results in a meta-analysis.

3. Briefly describe the **final sample of participants** after the recruitment and exclusion processes in terms of relevant characteristics (e.g., demographics, occupation, health status). These are the people whose data are reflected in the results of the study.
   Description: We want to document any characteristics reported about the final sample of participants that help us reason about the population they represent. This might be evidence that the sample is representative of the population we want to make an inference about or that the sample is not representative in some way.
   Location: Information about study participants will often be reported in the Methods or Results section.
   Importance: If the study participants are too different from the participants in other studies, it may not make sense to combine the results of these studies in a meta-analysis. If the study participants are too different from the population we are interested in making inferences about, the results from this study may not be applicable or relevant to our meta-analysis.

Measurement

2. How was the independent variable defined? Briefly describe how the author(s) implement the intervention and control conditions in the study.
   Description: Studies typically measure the effect of some manipulation or intervention. The specific condition which is manipulated is called an independent variable. An independent variable is assumed to have an impact on the outcome of interest (i.e., dependent variable).
   Location: Independent variables will often be mentioned throughout the paper but in particular in the Method and Results section.
   Importance: When conducting a meta-analysis, we will want to make sure that each of the studies we include measure the impact of the same independent variable. Otherwise, it doesn't make sense to

compare them. We also want to make sure that the manipulation of the independent variable produces conditions similar to the situation we are trying to make an inference about.

4. How was the dependent variable defined? Briefly describe how the author(s) measure or observe the outcome of interest in the study.
   Description: The dependent variable is how the authors measured the outcome of interest. This question asks what was actually done to get the measurements reported in the paper, and what set of conditions needed to be met in order for the authors to measure the outcome of interest. For example, the authors may have administered a survey, brought participants into the lab for data collection, or retrieved data that had already been collected for another study. In choosing how to collect measurements, the authors may qualify or limit what counts as an instance of the outcome of interest within the scope of their research.
   Location: This information should be reported in the Method or Results section.
   Importance: In order to compare results across studies, we want to know where the data came from and whether or not studies differ in their methods of measurement.

5. Was the outcome of interest dichotomous (i.e., an event that did or did not happen) or continuous (i.e., a scale with more than two possible values)?
   Description: This question asks about the data type of the outcome of interest. Studies with a dichotomous outcome often measure the rate of occurrence of some event (e.g., death, diagnosis, employment, etc.) and how the intervention of interest impacts this rate. In contrast, studies with a continuous outcome often seek to quantify the change in some scale or indicatory in response to the intervention of interest.
   Location: This information should be reported in the Method or Results section.
   Importance: Whether the outcome of interest is dichotomous or continuous determines what information we need to extract from the article in order to calculate the standardized effect size.

6. Was the measurement taken between-subjects such that there is only one observation per participant, or was the measurement repeated within-subjects over time such that there are multiple observations for each individual participant?
   Description: Some studies involve repeated measurements from individual participants over time. For example, authors may administer a survey to the same person at multiple points in time. Common reasons that authors may take repeated measurements within-subjects are: to compare outcomes before and after an intervention; to compare outcomes at different durations after an intervention; or to improve statistical power of their study design to detect patterns. These within-subjects measurements are in contrast with between-subjects measurements which each individual is measured only once.
   Location: Whether a measurement is within- or between-subjects will be reported in the Method or Results section.

Importance: Whether measurements are taken within- or between-subjects is important to consider when deciding whether to aggregate studies in meta-analysis. It is best practice not to combine studies that use repeated measures within individual participants with studies that make comparisons between individuals since these measurement strategies are thought to measure different things.

7. What was the time schedule for measurement in the study? When were measures taken?
   Description: Some study authors may stratify their data based on time. For example, they may report a measure taken at different durations after an intervention or baseline event. When authors stratify their data by time, it is often because the timing of events is important to their research question or the domain of study.
   Location: The timing of measurements will be reported in the Method or Results section.
   Importance: To the extent that the time course of measurements vary across studies, this may be important to consider when deciding whether to aggregate studies in meta-analysis. If the timing of measurements is dissimilar across studies, combining them in a meta-analysis will yield a result that averages over different time courses of measurement. At best this adds noise to the analysis; at worst it obfuscates the meaning of the aggregated effect size (e.g., if the effect is thought to change over time).

8. Do the authors report this measure separately for different groups of participants (e.g., grouping their sample by demographics or some other factor)?
   Description: Study authors may group or stratify their participants (e.g., by gender or age) and report effects within each group (e.g., presenting results in contingency tables). Authors create these groups by identifying factors or variables which are of interest in their analysis. Each of these factors has different levels or possible values it might take on. For example, for the categorical factor of gender, possible levels might be man, woman, or non-binary. For continuous factors like age, the study authors might create levels by binning their data into relevant age groups (e.g., 18-30, 30-50, 50-65, 65+) as a way of forming discrete levels. When authors present results for different groups, it is often because those grouping factors are thought to influence the effect of interest.
   Location: Grouping factors should be reported in the Results section, but the rationale for their importance may be located elsewhere or may not be explicitly stated.
   Importance: We need to decide whether these grouping factors are of interest for meta-analysis, or if we prefer to only document the overall effect. Even if we are interested in group-level effects, we will only be able to include them in our meta-analysis if they are reported across multiple studies.

Effect Size

Please use the following items to capture, in as much detail as possible, information reported about estimates of effect size. You will almost certainly find this information in the Results section. For each statistic reported about the effect of interest, please fill out the following items.

1. ~~What statistic is used to report the effect size estimate? This might be the same as the unit of measurement, or it might be derived from that measurement through a statistical model.~~
   ~~*Description: Effect size might be reported in any of the following ways:*~~
   - ~~*An average measurement in each group*~~
   - ~~*An average difference between groups*~~
   - ~~*A ratio describing the relative probability of the outcome of interest*~~
   - ~~*A correlation coefficient or R-squared value*~~
   - ~~*A regression coefficient*~~

   ~~*Location: Effect size should be reported in the Results section.*~~
   ~~*Importance: All of these different ways of reporting effect size represent the empirical results of a scientific study. They are estimates of the impact of the predictor or intervention of interest on the outcome of interest. These numbers are the primary information required for meta-analysis.*~~

4. How do the authors report the reliability of the estimated difference between the intervention and control groups? Please check one of the following.
   Description: Reliability is an indicator of how repeatable the estimation process is. Reliability statistics answer the question: How likely is it that if we repeated the same analysis process, it would produce a similar effect size estimate? Reliability might be reported as a test statistic such as a t-test or an ANOVA F-test. Alternatively, it might be a p-value, a standard error, or a confidence interval.
   Location: Reliability should be reported in the Results section alongside the effect size estimate.
   Importance: We want to document reliability information so that we know how uncertain the results from the study are. We need to incorporate this error into our meta-analysis.

5. If a statistical model produced the results, what covariates (if any) were adjusted for? These are factors which the authors controlled for when calculating effect size.
   Description: Covariates are variables other than the predictor and outcome of interest which are included in a statistical model. Including these other variables in the model adjusts the effect size estimate for the influence of these covariates. Specifically, the model splits the data into groups for each level of the covariates, calculates the effect of interest in each group, and takes a weighted average of the effect size across these groups. Covariates might be demographic variables like age or characteristics like health status which impact both the predictor and the outcome of interest. Most regression models in particular will have covariates, but other kinds of models may not.
   Location: The authors should report covariates for each model in their Methods or Results.
   Importance: Studies adjusting for different sets of covariates will estimate different effect sizes. It is best practice not to combine effects which are adjusted for different sets of covariates.