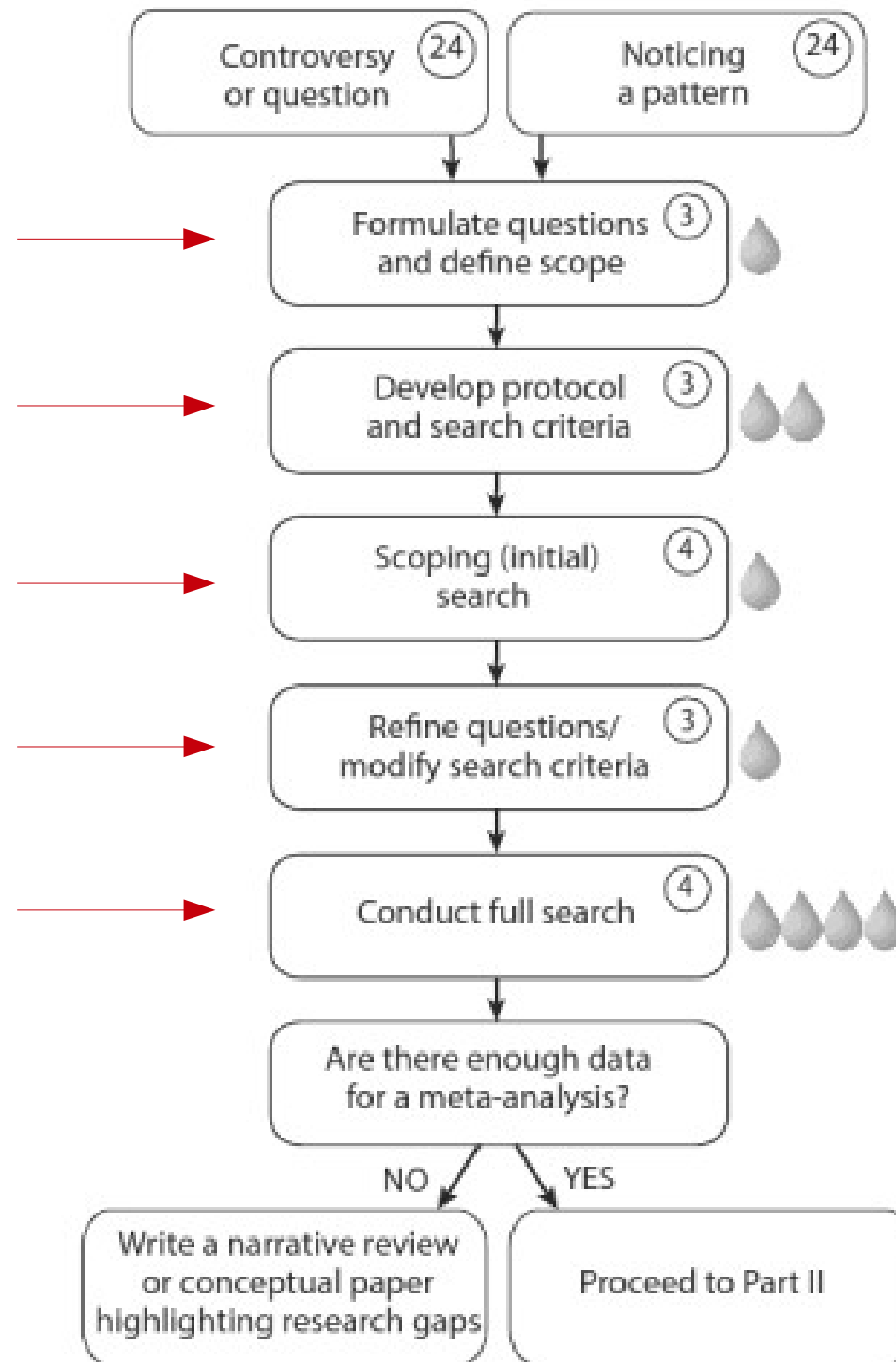


Finding and Gathering Data
Patrick Kearns
2/5/14

Outline

- (1) Defining a Question
- (2) Inclusion/exclusion criteria
- (3) Data collection
- (4) Data evaluation
- (5) Conducting the analysis
- (6) Data presentation
- (7) Interpretation of results



Defining a Question

- Need a sound question as to not introduce bias into the analysis.
- Without a solid Q, results are difficult to analyze and interpret post hoc.
- Heterogeneity between studies should not be considered important.
- Scoping for literature can help define whether a analysis is possible.
 - Get keywords to use in further searching.

Elements of a Question

- (1) Define a subject or population
- (2) Treatment/variable
- (3) Response
- (4) Control/Comparison group

Other question related items

- Study design of studies should be noted and tested for bias
- Define significance levels
- Subgroups related to Q can be incorporated if the data is discrete.

Study Inclusion

- Develop before performing search
- Can highly influence the outcome of the study
 - How?

Inherent bias associated with inclusion/exclusion

- Differs between people
- Good to have a second reader review the excluded studies (Kappa, pg 50)

Kappa

BOX 4.6. Kappa assessment: A worked example.

The table below shows the number of references accepted and rejected by two reviewers in relation to one another.

		Reviewer 2		
		Reject	Accept	Total
Reviewer 1	Reject	20	19	39
	Accept	1	110	111
	Total	21	120	150

Agreement expected by chance is calculated as: (row total × column total)/overall total, providing a second matrix, shown below.

		Reviewer 2		
		Reject by chance	Accept by chance	Total
Reviewer 1	Reject by chance	5.46	33.54	39
	Accept by chance	15.54	95.46	111
	Total	21	129	150

Actual agreement is calculated as the number of times reviewers agreed (i.e., both reject or both accept, from first matrix): $110 + 20 = 130$.

Similarly, expected agreement by chance is (from second matrix): $5.46 + 95.46 = 100.92$.

Kappa is used to calculate the agreement above and beyond the agreement expected by chance.

$$K = \frac{\text{actual agreement} - \text{expected agreement}}{\text{total number of trials} - \text{expected agreement}}$$

$$= \frac{(130 - 100.92)}{(150 - 100.92)} = \frac{29.08}{49.08} = 0.593$$

Kappa

BOX 4.6. Kappa assessment: A worked example.

The table below shows the number of references accepted and rejected by two reviewers in relation to one another.

		Reviewer 2		
		Reject	Accept	Total
Reviewer 1	Reject	20	19	39
	Accept	1	110	111
	Total	21	120	150

Agreement expected by chance is calculated as: (row total \times column total)/overall total, providing a second matrix, shown below.

		Reviewer 2		
		Reject by chance	Accept by chance	Total
Reviewer 1	Reject by chance	5.46	33.54	39
	Accept by chance	15.54	95.46	111
	Total	21	129	150

Actual agreement is calculated as the number of times reviewers agreed (i.e., both reject or both accept, from first matrix): $110 + 20 = 130$.

Similarly, expected agreement by chance is (from second matrix): $5.46 + 95.46 = 100.92$.
Kappa is used to calculate the agreement above and beyond the agreement expected by chance.

$$K = (\text{actual agreement} - \text{expected agreement}) / (\text{total number of trials} - \text{expected agreement})$$

$$= (130 - 100.92) / (150 - 100.92) = 29.08 / 49.08 = 0.593$$

Kappa

BOX 4.6. Kappa assessment: A worked example.

The table below shows the number of references accepted and rejected by two reviewers in relation to one another.

		Reviewer 2		
		Reject	Accept	Total
Reviewer 1	Reject	20	19	39
	Accept	1	110	111
	Total	21	120	150

Agreement expected by chance is calculated as: (row total \times column total)/overall total, providing a second matrix, shown below.

		Reviewer 2		
		Reject by chance	Accept by chance	Total
Reviewer 1	Reject by chance	5.46	33.54	39
	Accept by chance	15.54	95.46	111
	Total	21	129	150

Actual agreement is calculated as the number of times reviewers agreed (i.e., both reject or both accept, from first matrix): $110 + 20 = 130$.

Similarly, expected agreement by chance is (from second matrix): $5.46 + 95.46 = 100.92$.

Kappa is used to calculate the agreement above and beyond the agreement expected by chance.

$$K = (\text{actual agreement} - \text{expected agreement}) / (\text{total number of trials} - \text{expected agreement})$$

$$= (130 - 100.92) / (150 - 100.92) = 29.08 / 49.08 = 0.593$$

Kappa

$K = \frac{\text{actual agreement} - \text{expected agreement}}{\text{total \# trials} - \text{expected agreement}}$

$$K = \frac{130 - 100.92}{150 - 100.92} = 0.593$$

Inclusion To-do's

- Keep good records
- Its key to eliminate duplicate studies
- How do the studies relate to the scope of the analysis?

Data Collection

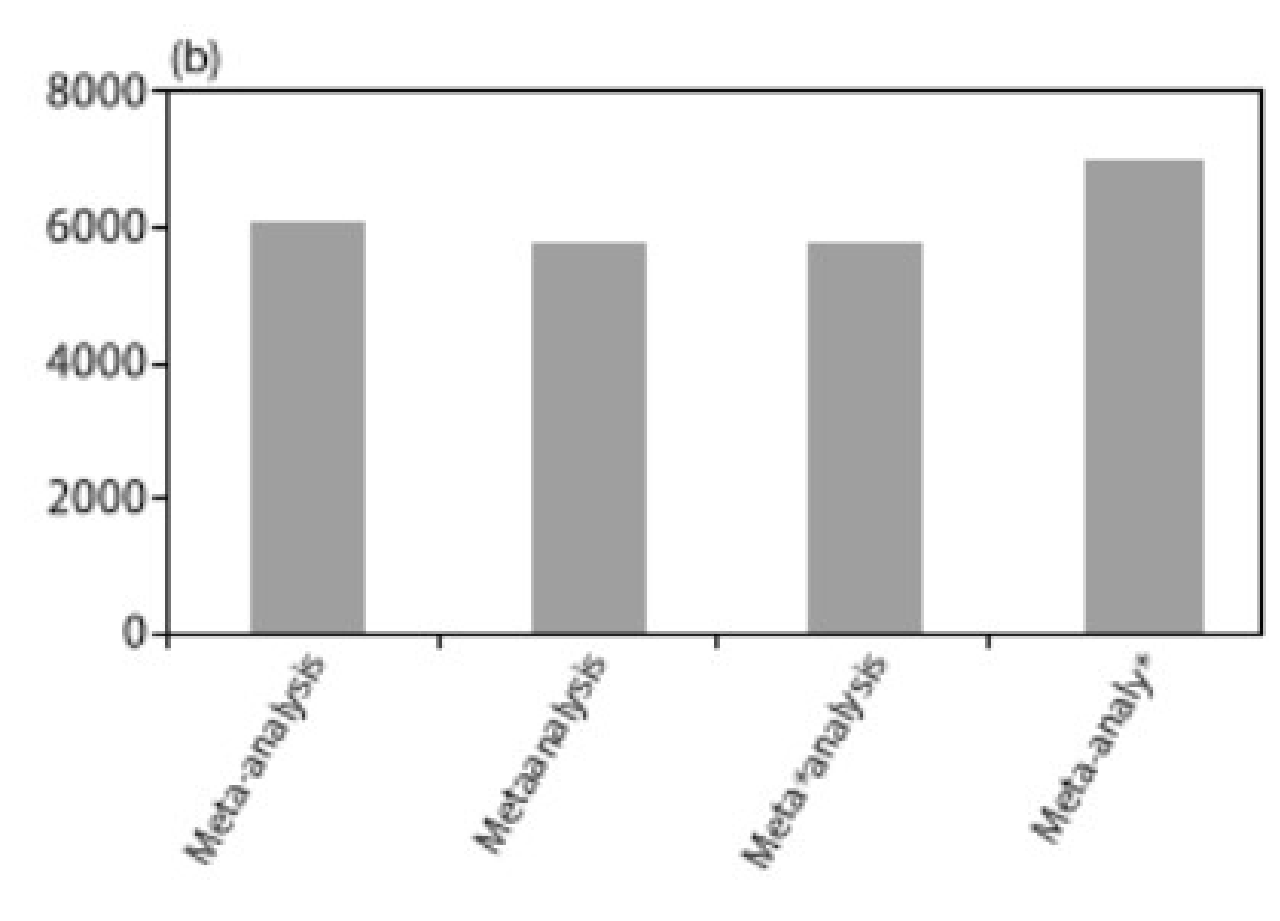


Search Criteria

- Need many sources
- Unwise to over represent a small number of organisms
 - Why is this bad?
 - Can it be avoided?
- Document each step of the way
- Can induce bias through poor methodology

Searching for Data

- Use key words from scoping
- Use modifiers to widen results pool (AND, OR, *, ?)
- *Is the question feasible or make sense?*
 - Redesign question



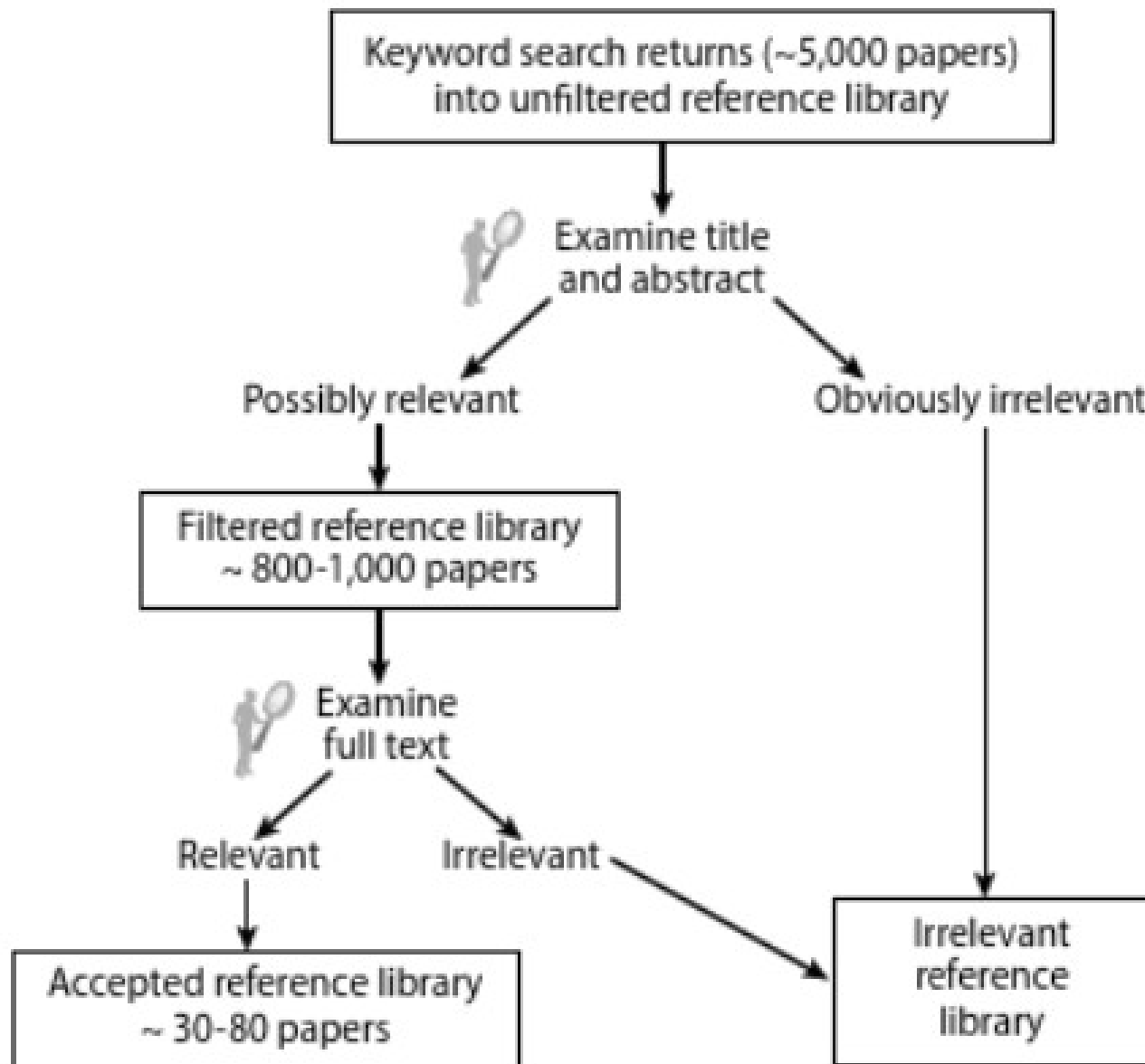
Searching for Data

- Use key words from scoping
- Use modifiers to widen results pool (AND, OR, *, ?)
- Is the question feasible or make sense?
 - Redesign question

How does environmental forcing affect microbial functions?

Data sources

- Published literature
 - Always good to check citation section
- Grey literature
 - *Book chaps.*, thesis's, gov. docs., bulletins, fact sheets, *conference proceedings*, posters
- Black literature
 - Field/lab notes



Data Evaluation

- Nonindependence
 - Factors influencing the outcome (e.g. caging)
 - Trying to test too many factors, they can be correlated
- Missing Data

Conducting the analysis

- Effect size
- Weighting
- Model choice
- Heterogeneity
- Missing data

Presenting data

- Similar to a research article
 - Introduction, Materials and Methods, Results, Discussion
 - M&M should allow another person to replicate your study
 - Results should include what groups/variable included as well as effect size, 95% CI

Data Interpretation

- How do your decisions during the analysis affect the outcome?
- Are the findings considered in the light of biological/practical significance?
- Highlight areas that are in need of analysis