Reading scientific literature

In this unit we will review the nature of science and how it is broadcasted through published, peer-reviewed literature.

1 Science: An approach to understanding truth

Science is using the collection of facts to understand truth. Facts are simply observations (e.g., measurements), and we usually refer to them as "data". Truth is simply the way things are (which is impossible to know with exactness because of error). Many fields of science use statistics to estimate truth- statistics are simply the tools (based on math) that we use to convert facts into estimates of truth. Statistics are powerful approaches that can be used to test hypotheses, which are proposed explanations of patterns observed in data. Don't confuse "predictions" with "hypotheses"; while predictions are based on hypotheses, a prediction is a specific foretelling of an outcome of a scientific investigation (e.g., if your hypothesis is that a specific gene causes a disease, a prediction would be to see variation in that gene among a sample of people where some have the disease and others do not). Theory is an explanation of patterns that has been corroborated by independent scientific investigations. When considering fact, hypothesis, and theory together, theory is at the top- they are the result of repeated hypothesis testing and are the closest thing to "proof" you can get in science. A scientific theory is very different than a idea based on little support. In other words, when someone says "only a theory" in science they are unaware of what a scientific theory actually is.

2 Introduction: What is scientific literature?

Knowledge obtained through the scientific method is disseminated through peer-reviewed publications, therefore being able to read and interpret scientific literature is a crucial skill for people in all fields of science. Today this is especially relevant, given that misinformation spreads rapidly within easy-to-read memes, videos, and blog posts. As a scientist, you are responsible for critically examining statements, reaching a conclusion yourself based on evidence, and sharing your conclusion and rationale. In order to do this appropriately, you must be able to read scientific publications.

Scientific papers are usually published in journals, which are periodic publications that contain a collection of scientific manuscripts. The lead author of a scientific paper is the individual who led the study. In many cases, the last author is usually the individual that supervised the study. All other authors are referred to as co-authors. The affiliations of all authors are typically located under the author line on a scientific paper.

The different sections of scientific papers typically include the following (but this varies by journal:

- **Abstract:** Brief summary of the manuscript
- **Introduction:** Provides the broad theory, background, knowledge gap, hypothesis, and sometimes predictions
- **Materials/Methods:** The approach for collecting and interpreting data
- **Results:** The outcomes of collection and analysis of the data (including statistics). Lists/summaries of the data in table/figure format.

- **Discussion/Conclusion:** Interpretation of the results in the context of the hypothesis and theory.
- **Acknowledgements/Funding:** Resources that helped the authors with the study.
- **References/Works Cited:** List of studies referenced within the manuscript body

You may ask yourself, "Why are primary scientific papers so hard to read? Why are they so full of jargon and extremely technical details? Why aren't they more straightforward?" Scientific papers are (or at least should be) full of all of the necessary material for the reader to (1) come to an informed conclusion [based on the background material, experimental design, data collection, and analyses] and (2) reproduce the study if desired. In order for these to aims to be met, generally the papers require some complexity. Secondary sources, such as review articles, newspaper articles, and blog posts, are generally more straight-forward. Yet these sources lack the details for minimally-biased conclusions to be reached. Secondary sources are quick ways to learn about the big-picture of studies, but in order to understand the science on a deep level the primary source must be examined. To this aim, we will be reading a scientific paper on an advanced topic in modern genetics.

3 How to read a scientific paper

"Scientific Paper" is a broad category that includes primary research articles (original research), review articles, case studies, and methods (i.e., descriptions of techniques and tools) papers. Just because something is a "scientific paper" doesn't make it legitimate- the reader is responsible to investigate whether a paper contains sound rationale and appropriate conclusions. Papers published in peer-reviewed journals are generally more dependable ("peer-review" is when experts in the field scrutinize a paper and write comments to the journal editor before the paper is accepted for publication), but even these can sometimes contain flaws (for example, see `https://blogs.scientificamerican.com/observations/controversial-caterpillar-evolution-study-formally-rebutted/`). As a scientist, you are duty-bound to be able to investigate scientific literature to validate claims. The world depends on people like you to confirm/reject claims that can have broad societal impacts!

Reading a scientific paper can be challenging, and your reading strategy will vary based on your objectives. For instance, if you are writing a review article that will be citing a paper or if you are presenting a paper, you will want to dedicate several hours to reading a paper to make sure you obtain a thorough understanding of the paper. In other scenarios you may just be looking briefly at a paper to understand the main takeaways. While the time you dedicate will vary based on your objectives, the general approach will be the same. Here is a helpful step-by-step process for reading scientific papers:

- Step 1: Read the Introduction section
  - Highlight terms that seem important; if you don't understand these, look them up and write down definitions next to where the words appear in the document
  - Identify the knowledge gap, hypothesis, and predictions
  - Generally the first paragraph and the last two paragraphs of the introduction will contain the most important information in the introduction
- Step 2: Skim the Methods section- focus on the experimental design

- Highlight sections about sample collection, data collection, and data analysis
- Draw out the experimental design (include the sample numbers, sample information, variables, and data collection
- Don't get bogged down by things you don't understand in this section- focus on the big picture

- Step 3: Look at the figures and interpret them with info written in the Results section
  - Take lots of notes on the figures themselves! You may think you understand one, but practice explaining everything about it to make sure
- Step 4: Read the Discussion and Conclusion sections
  - The first paragraph of the discussion usually provides a good review of the study objectives and execution. Use references to the figures that occur in the discussion to better understand the figures themselves. The Conclusion (or last paragraph of the discussion if there is no Conclusion section) provides the major takeaways for the study.
- Step 5: Read it again!
  - If you are seeking to obtain a thorough understanding of the paper, steps 1-4 will get you started- but you will need to read the paper again to fully grasp the study. A solid understanding of a scientific paper usually takes 2-3 hours of reading (at least).

Lastly, and most importantly, be patient with yourself. Reading scientific papers is HARD! Like most things, it takes a lot of practice. Don't think less of yourself if you feel like you are having a difficult time. Focus on what you know and stretch yourself to try to understand new concepts.

4 Activity 22:

Kyrou, K., Hammond, A.M., Galizi, R., Kranjc, N., Burt, A., Beaghton, A.K., Nolan, T., Cristanti, A. 2018. A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppresion in caged *Anopheles gambiae* mosquitos. Nature Communications 36(11):1062-1066. doi: 10.1038/nbt.4245

Cited above is a genetics publication by Kyros Kyrou et al. ("et al." is latin for "et alia". It is a common term used in academia when multiple people contributed to a work). The paper is available for download on Canvas under Files-¿ReadingMaterials. **Read this paper and answer the questions below.** Use the steps described above to help you read the manuscript efficiently. For many of these questions, you will likely need to look at resources outside of the paper itself (e.g., other papers, google searches, youtube videos, etc.). That is part of the process! Anytime you read a scientific paper, you should answer these questions.

Question 1: What is the journal?

Is this journal peer-reviewed? Is it well-regarded within the scientific community? The journal's impact factor can sometimes be informative of this (impact factor >1 usually means articles from the journal are cited somewhat frequently).

Question 2: Who are the authors?

Author order is informative. Generally the first author was the person who contributed the most to the study. The last author (sometimes called the "senior" author) was usually a chief contributor involved with the experimental design, funding, mentoring, and overall supervision. The second author down to the second-to-last author are ordered by their contribution to the study. Where are the first and last authors from? What stages in their careers were they at when this manuscript was published?

Question 3: Do the authors have any conflicts of interest? What are the funding sources? Do these introduce potential bias?

You should identify whether the authors have incentives to produce specific findings. For example, see a story about a recently redacted paper in a renowned journal (BMJ): `https://thehill.com/changing-america/well-being/prevention-cures/549679-paper-claiming-smoking-gives-protection-from`

Question 4: What is the primary question being addressed?

Is the study aiming to fill a knowledge gap? What is the knowledge gap?

Question 5: What is the hypothesis?

A hypothesis is a testable idea that explains an observed phenomenon. It is not a specific prediction from a specific experimental design.

Question 6: What is the experimental design?

Determining the experimental design can be very challenging- sometimes it is not spelled out succinctly within the paper. I recommend drawing out the experimental design to better understand it. While diagramming, answer the following questions: What study organism was used? What were the conditions the organisms were kept in / sampled from? What was the treatment? Was there a control group? What were the sample sizes for their treatment/control group? What data did they collect? How did they analyze the data? **See below for a description of "experimental design".** In many cases, it is easier to draw out an experimental design rather than describe it in sentence format.

Question 7: What did they predict?

What specific predictions did the authors expect from their experimental design? In other words, what patterns did they expect to see in the data analyses?

Question 8: What results did they see?

What patterns did they see in the data analyses? Do these match their predictions?

Question 9: What are the conclusions?

How do the results inform the hypothesis? Do they support the hypothesis? Why or why not?

5 The Scientific Method In Action

Understanding Experimental Design

Designing an experiment is the most important part of the scientific method. Even if a hypothesis is true, an experimental design that fails to appropriately test the hypothesis is useless. This can result in a waste of money, time, and other resources that can be devastating for a researcher. In some scenarios, published studies contain flawed experimental designs due to the failure of both the investigators and the reviewers to critically examine the study. As you may expect, this can be a source of misinformation and can be very damaging to the reputation of scientists. For that reason, it is important for all scientists to be able to interpret and design appropriate experiments.

So what is an experimental design? Simply stated, it is the overall procedure that will be used to collect and analyze the data. For an overview of experimental designs, see: `https://www.scribbr.com/methodology/experimental-design/`. Components that go into designing an experiment include:

1. How will the individuals in the experiment be collected?
2. What is/are the treatment(s) and how will they be administered? Is there a control group?
3. How are treatment/control groups being assigned to individuals? Is there randomization?
4. What will be the experimental conditions? In-lab or in-nature? Controlled experiment or observed study?
5. What will be the experimental timeline?
6. What kind of data will be collected?
7. How will data be collected?
8. How will data be analyzed?
9. What tools/equipment will be used for organism collection, housing, maintenance, data collection, and analyses?

As you read the Kyrou et al. paper to make sure you have a sound understanding of the experimental design. Then take time to evaluate the three figures in this paper. You'll notice that with a strong understanding of their experimental design it is much easier to understand the findings as presented in the figures.