The hitchhikers' guide to reading and writing health research

Arindam Basu[1]

[1]University of Canterbury

February 23, 2021

Abstract

In this paper, we introduce the concepts of critically reading research papers and writing of research proposals and reports. Research methods is a general term that includes the processes of observation of the world around the researcher, linking background knowledge with foreground questions, drafting a plan of collection of data and framing theories and hypotheses, testing the hypotheses, and finally, drafting or writing the research to evoke new knowledge. These processes vary with the themes and disciplines that the researcher engages in; nevertheless, common motifs can be found. In this paper, we propose three methods are interlinked: a deductive reasoning process where the structure of the thought can be captured critically; an inductive reasoning method where the researcher can appraise and generate generalisable ideas from observations of the world; and finally, abductive reasoning method where the world can be explained or the phenomena observed can be explained or be accounted for. This step or reasoning is also about framing theories, testing and challenging established knowledge or finding best theories and how theories best fit the observations. We start with a discussion of the different types of statements that one can come across in any scholarly literature or even in lay or semi-serious literature, appraise them, and identify arguments from non-arguments, and explanations from non-explanations. Then we outline three strategies to appraise and identify reasonings in explanations and arguments. We end with a discussion on how to draft a research proposal and a reading/archiving strategy of research.

Introduction

Research methods is a general term that includes several related activities: (1) observing the world around you, (2) critical reading of literature or critical appraisal of the video or lectures you come across, (3) evaluation of the information, (3) writing a research proposal, (4) drafting a research protocol, (5) mode of collection of information relevant for your research, (6) analysis of data, and finally (7) writing the actual research piece that will evoke further studies. In this tutorial, we are going to discuss how you can critically read and write research. This is a fundamental skill for every researcher in any discipline. We will first lay out the general principles and then we will discuss the more specific issues.

How to read an article or a report

Six types of statements

As you read an article, pay attention to each sentence and how sentences are linked; also pay attention to how paragraphs together hold the entire paper and flow from a discussion of the background of the research,

the methods the authors adopted to test or find the key findings, the description of the findings themselves, and finally, the discussion and conclusion. This format, termed IMRAD (introduction, methods, results, and discussion) has standardised the practice of writing research papers and we are not going to elaborate a discussion of this format here. For more information on how to use this format, consult a relevant journal where you want to send your paper. Todorovic (2003) has provided a tutorial on how to use IMRAD format (Todorović, 2003). Taking a modular approach to reading a paper, we recommend that you note each sentence carefully as you read a paper. Tom Chatfield (2017) describes in his work, "Critical Thinking" that you can classify each sentence/paragraph that does not include any reasoning embedded into one of the following types (Chatfield, 2017):

- **A description**. – For example, "PM10, Ozone, Carbon dioxide, oxides of sulphur and nitrogen are referred to as criteria pollutants". When something like that appears in a passage, if it is not accompanied by a reason, we label these sentences as descriptions of state of things. These are not accompanied by any reason behind such statement.
- **A summary**. – A summary is usually presented in the form of an abstract, but they can also be found in any part of the passage. For example, common places where summaries appear are methods section. Example, "We conducted a case-control study design with 100 cases and 100 controls that were matched for age and gender. We provided each case and control a questionnaire about their dietary intake and we also measured their portions using cups and spoons". Note that when you read a passage such as this, the authors have provided a string of statements but they have not provided any reason to back up these statements. These are referred to as summaries.
- **An opinion**. – An opinion is a first person perspective of facts or statement that expresses one's personal position on a topic. A description is a third person perspective where a the narrator would state something happened or something was being done, but usually does not post his or her own position, here the situation is different. Example: "We think use of electric vehicles can reduce air pollution in major cities." Often, the opinions and other descriptions are in the form of questions or suggestions.
- **A belief**. – A belief is similar to opinion except in belief, the expression of idea is expressed with a moral undertone. For example, "I believe for the sake of saving the planet's health, and for environment and climate, people should adopt a more vegetarian oriented eating and lifestyle" (this is not an actual quote but an illustration (see below) to highlight the idea of how beliefs are stated. So, when you see a statement expressed in a piece of literature that states a moral undertone and without any reasoning of such a statement, label it as a "belief" statement.
- **Clarification**. – A clarification is a statement that relies on another statement that has preceded it, and a clarification statement would be making some claims explicit or easier to interpret. For example, "Criteria pollutants, that is concentrations of PM10, PM2.5, Ozone, Carbon dioxide, oxides of nitrogen and sulphur, are considered to be air pollutants that have significant health effects". Note that in this statement too, the author has not expressed any reason as to why they are "criteria" pollutants or what criteria they fulfill or why they are labelled as such. All he has done is to state in the first place and then add an elaboration.
- **Illustration**. – Illustrations are similar to clarifications, except in illustrations, the author adds a specific instance or an example to portray the idea that he is writing about. For example, "Exposure to inorganic arsenic results in skin lesions - dark and light spots next to each other on the trunk and hands and feet". Here, dark and light spots further qualify the "skin lesions" although they do not add anything more by way of a reason as to what is the rationale of such a statement.

So, description, summary, opinions, beliefs, clarification, and illustration provide a class of expressions and statements that you should note with the caution that they are not associated with any reasoning. There are places and situations to use them, as descriptions and summaries contain data that are essential for developing further arguments.

You should be cautious about opinions, beliefs and rhetorical manipulations. In general, treat opinions and beliefs with circumspect if not associated with reasoning; they may not be useful for developing arguments or explanations which you will seeking as you read or be ready to learn from academic writing. Besides, you should be cautious about rhetorical manipulations and rhetorical elements. Rhetorical elements are statements associated with "emotional languages". For example, "It is high time we took action against the fracking industry as they are playing with the lives of millions and the life of the planet". Note, that in statements such as this, we do not see an associated rationale or reasoning that we can argue with or use as explanation. Irrespective of their intentions, you should treat emotional laden language that are not associated with any reasoning with skepticism as well (note that this statement itself is an opinion).

This leaves us with two other classes of statements/passages/paragraphs that you will come across in any academic or scholarly writing or "paper": explanations and arguments. Explanations are statements that rely upon preceding statements that are either descriptions or facts, and explanations build upon them by adding reasoning statements. However, explanations only offer reasoning behind facts or state if a pattern is being observed or some form of generalisation. For example, "Compared with those who did not have breast cancer, those who had diagnoses of breast cancer were 1.56 times likely to consume high cholesterol diet. This suggests that high cholesterol diet on a regular basis is a risk factor for breast cancer." This statement has two sentences, and the second sentence is a "conclusion" that high cholesterol diet consumption on a regular basis is a risk factor for breast cancer; the first statement is a rationale for arriving at such a conclusion that those who did not have breast cancer were found to consume less cholesterol in their diet. When we have statements such as this, we label these as explanations. Explanations are some of the most common statement sets in scholarly literature and we should pay attention to the various facts and statements that make up explanations.

Now, consider another class of statements: "Compared with those who did not have breast cancer, those who had diagnoses of breast cancer were 1.56 times likely to consume high cholesterol diet. This suggests that high cholesterol diet on a regular basis is a risk factor for breast cancer. Hence we recommend that cholesterol contents of diet should be restricted". This statement or set of statements too, have pointed to facts or descriptions, they have provided rationale or added reasonings, and over and above, they have provided a "call to action" or some form of persuasion, as to why is this explanation contextual. When you have a series of statements that are not only associated with reasons, but also associated with "a call to action" or some form of persuasion, then you deal with "arguments". Arguments, as you may see, are the ones with which you can engage actively as you read literature and you will provide counter arguments or provide alternative explanations to counter their claims. We will deal with arguments and explanations in the rest of this paper.

How to construct a standard form of argument and why will you do so?

The first thing you do when you read a paper is to scan for these various forms of sentences or statements and look for or scan for arguments that the authors have provided. If you look for arguments and explanations, then the authors would have provided you with a final conclusion or a final concluding statement, and several supporting statements to support the conclusion as reasoning. The statements that the authors have provided as reasoning to support their claim for hte conclusion they have made in the paper are referred to as "premises". We will differentiate between two different classes of premises: explicit premises, and implicit premises. Explicit premises are those statements that act as premises and those that hte autors have articulated in the paper. You can easily identify them as they are associated with the conclusions. You spot conclusions in various sections of the paper by noting identifier words or phrases. Look for phrases such as "hence", "given", "given that", "therefore", and so on; these words and phrases are those that signal that

3

a qualifying statement is going to follow, and a set of statements that have supported the statement will be found either next to the concluding statement or preceded the conclusion. Look for these statements. A paper is based on more than one argument and explanation. Each argument will have exactly one final conclusion, and several premises.

Compared with an explicit premise, which you will find in the article, you will need to supply the implicit premises by closely reading the explicit premises and your own common sense. For example, consider the following statement:

"Since the beginning of the agricultural and industrial era, humans have generated greenhouse gases and emitted the gases in the atmosphere. As a result, the global temperature has increased resulting in global warming".

If you read closely, you sense that this statement is actually made of three separate sentences:

- Explicit Premise: Since the beginning of agricultural and industrial era, humans have generated greenhouse gases and emitted them to atmosphere (this is an explicit premise because the authors have stated it so)
- Implicit Premise: Greenhouse gases cause increase in global temperature (this is an implicit premise because the authors have not stated this, but if you supply this premise in between, then the following sentence becomes clearer)
- Conclusion: Therefore (shorthand form of "as a result"), the temperature of the atmosphere has increased

Now this example shows you how a standard format of argument is constructed. You start with a conclusion, and there is only conclusion per argument, and the conclusion of an argument can well become a premise in a following argument that is connected to this argument and so the paper flows as a series of connected arguments or independent arguments that act as premises of each other resulting in a final conclusion of the paper. But each argument will have only one conclusion. You always start with a conclusion and you always first identify the more obvious, explicit premises. Then, after you have identified all explicit premises that lead to an argument, you fill in the implicit premises that lie in between the explicit premises and the conclusion. In the process, you may see that you have supplied some intermediate conclusions as well and that is OK. For example, in the statement above, where we supply the implicit premise "Greenshouse gases cause rise in temperature", could well be an intermediate conclusion. The implicit premises are also referred to as "implicit assumptions" or you can call them "assumptions".

Why are we breaking our heads over these things? Are these not common and can we not always assume them? The answer is "No we cannot". We are not allowed to go beyond what the authors have stated or supplied as facts, BUT, we can always argue by linking these explicit assumptions with other reasonings that either follow from them or are implied by them, which we term as "implicit assumptions". The implicit assumptions are the ones that we will need to examine. For instance, here, we would need to verfiy or justify that the claim, "greenhouse gases, when they accumulate in the atmosphere, will actually lead to increased temperature". It turns out that this fact is "absolutely correct" or "certain" as this "proof" is obtained from physical experiments where Svante Arrhenius in the 1700s showed that gases such as carbon dioxide when they accumulate in closed systems increase the temperature of these systems. So there is a rational basis of this argument and we can accept it. But it will not always be easy or simple as this. This is also an opportunity where you can argue the "missing" pieces in an argument that then forms the basis of your own research.

4

So, here are the rules of developing standard format arguments (follow these steps, we will see in an actual reading piece):

1. First, identify the strongest argument you can find in a paper or a piece of writing. Also, assume that the author is authentic and is knowledgeable to write what he or she is writing about. This is referred to as the principle of charity. While we label this as principle of charity, it has nothing to do with philosophically being kind to the authors or adhering to anyone's authority. It is done for a practical purpose. If you do not adhere to the principles of charity, and if you do not adhere to the strongest argument that you can examine, then you cannot claim that a statement or thesis is to be refuted, as referring to a weak argument makes no sense. Hence, you should always identify the strongest argument and provide the author the benefit of doubt.

2. Second, you should always identify the conclusion of an argument. Start with this first. Set up a spreadsheet and enter this as the first element in a cell. You will need to add cells to the spreadsheet on the TOP of this cell to enter explicit and implicit arguments, but do this first.

3. Third, write the explicit arguments clearly. Identify ALL explicit statements or arguments that the author or authors have stated and write them clearly on the top of the "final conclusion". So, the final conclusion will lie at the bottom and on the top of the final conclusion, you will list all the premises that are explicit. At this stage, do not worry about the order in which you identify them or write them. It is important that you should not miss any of the explicit claims. If you are in doubt, include the explicit premise. You can take it out later.

4. Fourth, mindfully read the explicit premises and try to identify what implicit assumptions or implicit premises they are based on. Some of these assumptions will be intermediate conclusions, and that is OK. At this stage, you are only going to identify the implicit assumptions or implicit premises.

5. Now that you have identified ALL the explicit and the implicit premises of the argument as well as the conclusion (final conclusion) of the argument that you are about to examine, read the premises mindfully. Do you see that these premises are linked or unlinked? Linked premises are those premises that should depend on one another to make sense or lead to the final conclusion. For example, consider the following statement: "We found increased concentration of inorganic arsenic in the groundwater of the Region X where we sampled the water. Experimental studies suggest that exposure to high concentration of inorganic arsenic in the groundwater leads to bladder cancer among mammals that were studied. We suggest that we should screen for evidence of bladder cancer among residents in Region X". Let's construct a standard format argument from this (EPx == Explicit premise x where x = 1,2, 3, etc; IPx == Implicit premise or assumptions x where x = 1,2,3, ...; ICx = intermediate conclusion x where x = 1,2,3, ...; FC = conclusion):

- EP1: We sampled the groundwater in Region X
- IP1: We are confident that our samples are representative of all groundwater in Region X
- EP2: We found in our samples of groundwater in Region X high concentration of inorganic arsenic
- IC1: From here we conclude that groundwater in Region X contains high concentration of inorganic arsenic
- EP3: Exposure to high concentration of inorganic arsenic in the groundwater causes bladder cancer among mammals;
- IP3: Human beings are mammals
- IP4: Inorganic arsenic in the body of all mammals act similarly
- IC2: Therefore human beings, when exposed to high concentrations of inorganic arsenic will develop bladder cancer
- IP5: Bladder cancer can be detected in subclinical stages (that is, stages of cancer when the cells are small enough and have not resulted in clinical manifestations of the disease)
- IP6: If they can be detected early enough in the disease cycle, then effective cures are available
- FC: We should screen for evidence of bladder cancer among residents of Region X

Can you see the complexity of information hidden in a simple four sentence statement? In the statement, we

have stated that we have found high concentration of inorganic arsenic in the groundwater of region X and that we know that if certain mammals are exposed to high concentration of inorganic arsenic, then they are likely to develop bladder cancer; therefore, we recommend that we should search for people who have not yet developed signs of bladder cancer but can develop and start a programme. Each of these implicit assumptions and intermediate conclusions are open to further investigations and investigations as to whether there are knowledge gaps. For example, where is the evidence that inorganic arsenic acts similarly in the body of all mammals and all mammalian systems are sufficiently identical so that we can extrapolate the results from one species to another. Are there inter-species variation in which inorganic arsenic is metabolised? Is it only inorganic arsenic or are there other compounds that are either associated with inorganic arsenic or their byproducts that can give rise to cancer that we should be mindful of? Besides, are these premises linked to one another? For example IP3 and IP4 (human beings are mammals and inorganic arsenic acts similarly in the body of all mammals similarly) are linked to each other and are crucial for reaching the final conclusion. If one of them fails, the final conclusion will fail. Similarly, IP5 and IP6 are linked to each other. Although other premises may not be as linked and can be treated as unlinked premises and they stand on their own.

So, we suggest that when you read a piece of literature critically, this is the first thing you should do. Start noting the types of expressions, and note the sentences or statements that are associated with some form of reasoning with them. Note which of them are explanations so that they do not call for any form of "call to action" or "persuasion", and note which ones call for action and are aimed to persuade you to take certain action. The ones that aim to persuade you in some ways are 'arguments'. Then, once you have noticed arguments & explanations, you start constructing standard formats to develop explicit and implicit premises, and lay out your argument maps. After this, we will examine the nature of such reasonings. Are these robust or sound arguments? Are these cogent arguments? Are these arguments forceful enough? Can we explain these relationships and if we can explain the relationships and would like to answer the question, "why we see what we see, what can we do about them?" This is where classifying and analysing the nature of the reasoning makes sense.

Logic, probability, and explanation: the three classes of reasonings

After you have constructed the standard form of argument, you identify the type of reasoning that the authors have used to arrive at their conclusions. Do these conclusions make sense to you? We state that there are three classes of reasoning that you should be aware of: reasoning that are pure logic and express relationship between implicit and explicit premises and the conclusions they reach purely by the force of logical thinking. In doing so, you rely on your common sense and your close reading of the statements. This class of reasoning is referred to as "reasoning by deductive logic". Reasoning by deductive logic may sound like detective novels or detective work and indeed they are like that. All we say here are that the statements should make good logical sense, or good common sense to us.

Reasoning by deductive logic: the play of logic

Consider the statement, "Exposure to high concentration of inorganic arsenic through water in all mammals result in bladder cancer. As humans are mammals, when human beings are exposed to high concentrations of inorganic arsenic through drinking water, they will develop bladder cancer". If we break up this statement into a standard form, we get the following:

- P1: All mammals, when exposed to high concentration of inorganic arsenic in drinking water, will develop bladder cancer … (1)

6

- P2: Humans are mammals ... (2)
- C: Therefore, humans, exposed to high concentration of inorganic arsenic in drinking water will develop bladder cancer ... (3)

Now each of these sentences are true. This argument is a deductive argument and note the format:

All humans are mortal.

I am a human.

Therefore I am mortal.

Note also that when we discuss something like this, or a logical set of statements, we also state that there is a certainty that will happen. For example, there is certainty that if mammals are exposed to inorganic arsenic through drinking water or water they drink, then they will develop bladder cancer, and as humans are mammals, this holds true for them as well. This element of certainty differentiates reasoning by deductive logic from the reasoning by inductive logic, we will discuss next.

These things may appear intuitive, but we need to keep in mind a few other rules as well that will become handy as we examine text where these rules are often violated as people write and as people think. Note that everything we write here not only relates to what we read, but also what we see in the form of video or graphics, what we hear when people speak as in lectures or in everyday language, or what we experience through our thinking. Here then are a few elements of this logic:

If X then Y ... (1); here, in this structure we term X as antecedent, and Y as consequent

X is true: therefore Y is true (this is referred to as affirmation of the antecedent)

Example: If X is a mammal, then exposure to inorganic arsenic through drinking water will cause bladder cancer; X is a mammal, therefore it is certain that X will develop bladder cancer exposed to arsenic through drinking water.

Here, this argument is a "valid" argument.

Note another one:

If X then Y ... (1)

Y is true: therefore X must be true .... (2) (this is an invalid argument and is referred to as affirmation of the consequent)

Example: "if the surface temperature of the seas increase, then there will be increased intensity of cyclones when they have landfall; we have experienced increased intensity of cyclones during landfall; it follows that the surface temperature of the sea have increased"

This is an invalid argument, even if both the statements are true (that is, there have been increase in the cyclone intensity over the past years, and that the sea temperature has remained an all time high). But just because the cyclone intensity has increased does not imply that the sea surface tempeature is high. The increase in the intensity of cyclones can be due to many other factors. You will frequently encounter these sort of errors in emotion laden languages and in non-fiction written to influence or present specific points of view. For example, in a recent book on Climate change, an author has claimed that he has experienced severe storms in Mumbai, India when he was young; remembering the storm surge, he now states, that the storm surge indicates the play of climate change. This is an example of "affirmation of the consequent" and is an invalid argument. What he argues is something like this:

7

If the global temperature increases, there will be severe cyclones and their intensity will increase ... (1)

I have experienced heavy cyclones or high intensity cyclones

This implies that global temperature have increased

It might as well be that the global temperatures have risen and also that storm intensity or cyclone intensity have been high, but the way he presented it do not make logical sense. Such logical leaps are invalid logical leaps and you must keep your eyes and ears open to catch this type of logical inconsistency.

Consider a third case:

If X then Y .... (1)

Y is not true; therefore X is not true as well .... (2) (This is valid argument and is referred to as negation of the consequent)

Example: "If there are sudden increase in the concentration of PM2.5 in the air, there will be abnormally heavy cases of deaths due to heart attacks in the city; deaths due to heart attacks have been steady or at least not in record high levels: hence, the sudden unexpected rise in PM2.5 has not happened". This statement may be convoluted and we will see in the next section that these things fit well with reasoning by induction, but this illustrates the situation with negation of the consequent that this is a valid argument. The fact that Y is not true signals that X did not happen if X was the only reason Y should have happened.

But, on the other hand,

If X then Y ... (1)

X is not true: therefore Y is not true as well ... (2) is invalid and is referred to as negation of the antecedent.

If we were to continue with the previous example, it would be stated something like as follows:

"Sudden increases in PM2.5 levels will cause high rates of admission due to heart attacks in emergency rooms; PM2.5 levels have not risen high today, therefore we will not expect high rates of admission due to heart attacks in emergency rooms". This assertion is not true as there may be other reasons why there would be high case loads of heart attack related hospitalisations.

We have here scraped the surface of what may happen with logic and how you should keep your eyes and ears open to the possibility of how people play with logical conclusions. However, in health sciences, these errors are relatively rare because the scope of deductive reasoning is not very popular. Instead, in health sciences and health care we are interested in another form of reasoning; this is reasoning by induction.

Sound and unsound arguments

In the above sections, we dealt at length about validity of the logic and conclusions we have drawn from the premises. However, along with the fact that the logical structure of the arguments be valid, the statements themselves should also be true. If the arguments are not valid in themselves, then the question of evaluating whether the statements themselves are true or not does not arise. Consider the following argument:

If an alien spacecraft with yellow colour hovers over the North Island of New Zealand, one person will prematurely die

We have not heard about premature deaths so far; therefore an alien spacecraft with yellow colour has not hovered over New Zealand. (This is taken as a spoof on a theme from "The Hitchhikers' guide to the galaxy")

Well, we cannot find fault with the logical structure of the statements, but the statements do not make sense. This sort of a situation, where either the logical structure is false, or the sentences do not make sense, are referred to as unsound arguments. Sound arguments are those arguments where you get both logical sense & that the facts are true as stated.

Reasoning by induction: the case of probability & sampling

In reasoning by induction, we also set up premises and conclusions but the premises and conclusions take the shape where premises are facts or observed phenomena. The observed phenomena from specific examples or specific instances then lead to generalised statements. The generalised statements are not absolute statements, we express the generalised statements using probabilistic language. Consider the following example:

P1: In our study, we found that those who had lung cancers were twice as likely to be exposed to environmental tobacco smoke than those who did not have lung cancer.

P2: Other investigators have noticed similar associations with other respiratory diseases as well.

C: We conclude that exposure to environmental tobacco smoke is a likely risk factor for lung diseases

Such statements are commonly expressed in the discussion or introduction section of studies, or theses. What's going on here? In the first two sentences (P1 and P2), the authors describe two specific observations. Then in the third sentence, "C", they make a generalisable statement, that based on P1 and P2, they can conclude that what they observed can be generalised to a wider truth (that is exposure to environmental tobacco smoke is a risk factor for lung disease). But in stating that, they do not write in absolute terms; instead they mention certain levels of uncertainty in the concluding sentence. This is the essence of inferential logic and this is the essence of inductive logic: we start with observations of the particular truth and from there, we extend to the larger or more generalisable truth. But because we are limited to specific observations, therefore, we cannot make statements in absolute terms. Consider the following statements:

P1: For each of the last ten years, we have experienced, using the averaged annual temperature trends, each subsequent year has been warmer than the previous one;

C: Hence, it is likely that the next year will be hotter than this year.

Here, we see that there has been an observation and then based on the trends observed over the past, there is a statement that it is likely that the next year will be hotter than the previous year. But how likely? This is a good question and we will reserve the meaning of this term till the next example. Consider the same or similar statement as above but stated in a different manner:

P1: Last year was hotter than the year before.

C: Hence, it is likely that the next year will be hotter than this year.

Again, the same question: if it is likely, how much faith to be have in how likely?

This is where we talk about probability and sampling, and eventually, we will discuss about probability estimates. The argument here can be either a strong argument (as in the first case) or a weak argument (as in the second case). What is going on? The next year may jolly well be hotter than this year, but if we were to make that statement only on the basis of one instance (that is this year has been hotter than last year and hence next year is going to be hotter than this year), we cannot be too certain. The probability will indeed be very low. This is the reason why case studies based on N = 1, or anecdotal evidences are not very forceful or cogent as arguments. On the other hand, in the first example, we see the same argument is being made, but in this case, we see that we argue on the basis of the fact that we have ten years' worth of data on which we bank on. This makes our probability estimate more likely than the case with only one year worth of experience.

Let's illustrate this with the following example:

(Impossible, probability = 0) .... [0.5 50/50 chance] ....
(Certain, probability = 1.0)

The above scenario generates three probability of a certain event, for example, the chance of rain tomorrow (say). Either, you can state that there is no chance or probability of rain tomorrow, in which case you will assign the probability of rain referred to as p(Rain) "0". Or, you are not sure at all, and you claim that the probability of rain is 50%, that is a 50% chance that it is going to rain and a 50% chance it is not going to rain, so we assign it the probability of rain or p(Rain) = 0.5; on the other hand, if it looks like that it is definitely going to rain tomorrow, then you state that the probability of rain tomorrow is 100%, and in terms of probability language we write, p(Rain) = 1.0; you can see from here that probability is "bounded", that is it can be at one end 0 and at the other end 1. What if we were able to free up so that the "chance" of something occurring versus not occurring would have a boundary of 0 at one end and infinity at the other end? This is useful for some estimation issues as we will see, and we call this likelihood, so let's explain that now. We use the same diagram as before but this time we use a ratio and we call this ratio "likelihood or odds" (we will visit Odds when we will discuss study designs such as case control study designs). So, here we go again with the rain example:

(Impossible, probability = 0) .... [0.5 50/50 chance] ....
(Certain, probability = 1.0)

Now we introduce the ratio of something happening versus something not happening. That measure is referred to as likelihood or Odds. So, in case where we stated that it was not going to rain at all, we express this in terms of probability of rain or p(rain) = 0, and probability of no rain or p(no rain) = 1.0. Then, the odds of rain as opposed to no rain tomorrow would be expressed in terms of p(rain) / p(no rain) and you can see we are dividing 0/1.0, and so the odds(rain) = 0. What happens when we have 50/50 chance of rain. Here p(rain) = 0.5 (50 out of 100), and equally p(no rain) = 50/100 or 0.5. Therefore here the odds(rain) = 0.5/0.5 = 1. In the extreme case, where we say that it is definitely going to rain, the p(rain)/(1 - p(rain)) will be infinity as we will divide 1.0 by 0 and it is not estimable and given a value of inf. So, now we have this:

(Impossible, odds = 0) .... (fifty/fifty chance, odds = 1.0) ....
(Certain 100% chance, Odds ~ Inf)

So that is the concept of probability and odds that probability, and remember some rules (call them proba-

bility sutra):

- p(X) will lie between 0 and 1, it cannot be less than 0 (that is, there no such thing as negative probability) and it cannot be greater than 1.
- If there are mutually independent events that makes up our universe, let's say X and Y, then p(X) + p(Y) = 1; what this means is let's say you want to toss a coin. As you toss a coin, there can be only two outcomes, heads or tails. We will say p(head) + p(tail) = 1. If the probability of head and tail are equal then you know that p(head) = 0.5, and p(tail) = 0.5 and so together they are 1.0; can you now work out what it will be with the rain thing? p(it will rain) + p(it will not rain) = 1.0; you can vary the probabilities of raining versus not raining, they will always add up to one as these are mutually exclusive events.
- A third rule is this: if X and Y are "disjointed or mutually exclusive or independent events", then their joint probability of occurrence will be p(X)*p(Y). An example is to think in terms of tossing a fair coin. If you toss a fair coin, the probability of heads (shorthand:"H"), p(H) = 0.5 (you can see that the probability of tails or p(T) = 0.5 too). The first time you toss a coin (we call it number of trials as N, each act of tossing a coin is referred to as a trial), the probability of head is p(H) = 0.5. The second time you toss a coin, for that toss, the probability of head is still 0.5 (as these are two independent events). What is the chance that this time too, you will have a probability of head? This is where the third rule comes into play: as these are independent of each other, we will say that you will two heads in a row, say something like p(HH) = 0.5 * 0.5 = 0.25.

You can probably guess where I am getting with this. For any random event (such as a coin toss, or say something like the weather or rainfall), where there is no pattern, you will be able to multiply their individual probability and be able to predict what will the probability of an event happening over and over again (or finding a pattern) under conditions of random occurrence: just multiply their probabilities and you will get the value. So, if we were to guess say that the probability of each succeding year hotter than the previous one was 0.5, then if you were to base your prediction on the basis of just one year's experience (N = 1), you'd say p(next year hotter than this year) = 0.5; the odds will be 1.0, and we are none the wiser. Under the conditions of randomness, what is the probability that each succeeding year will be hotter than the previous year? (again assuming that these things are truly random or 50/50 chance), you can construct a chart like this:

Event	Individual Probability	N	Joint Probability
Year hotter than last	0.5	1	The first time, so 0.5
Year hotter than last	0.5	2	0.5 * 0.5 = 0.25
Year hotter than last	0.5	3	0.5 ^ 3 = 0.125
Year hotter than last	0.5	5	0.5 ^ 5 = 0.031
Year hotter than last	0.5	10	0.5 ^ 10 = 0.0009

Let's explain the above table. The first time, or if you were to think in terms of the probability of the next year being hotter than this year based on only one event (N = 1) where say this year was hotter than the last year, you would say there is a 50/50 chance, so probability will be 0.5. What is the chance that five years in a row, you will see that the succeeding year is hotter than the previous one? Fairly low, as you can see it turns out to be 3 percent. So, 97% chance that such events would not occur if these events were not systematic or random as we would expect. How about ten years in a row? That probability is really low, like 0.009 % or less than even 0.01%, or one in a thousand chance. But that is only under the conditions of randomness. If that does not happen like that, there must be something systematic going on and you will

be most likely right to conclude that the next year is likely to be hotter than this year as well.

This brings us to the concept of sample size and sampling. You saw that if you were to base your decision on the basis of a single sample (such as just one year of data), then your chance of prediction is much lower than if you were to base your predictions on the basis of a lot of trials or a lot of years of experience. This is the issue of estimation. If you want to estimate an effect, or say if you want to study the prevalence of a condition (how much of a disease prevails in the society), then you need large enough samples of data. But how much is too much and how much is too little. This is why you conduct sample size estimation. There are several calculators that you can use online, and you can use sample size procedures in statistical programming environment such as stata or R. You can use the free open sourced web based sample size calculator at openepi.com(men) here:

http://www.openepi.com/Menu/OE_Menu.htm

Sampling refers to the process and steps of conducting a process where you select a representative set of data that resemble as close as possible the original population. You run all your estimations on the sample and then "extrapolate" your conclusions as to what may happen in the population. When you do sampling, you will need to keep in mind that your sample must have participants whose distributions should be as close as possible to the population. Different investigators, depending on the specific research question, use or adopt different strategies to select the samples of their studies. Some investigators use simple random sampling, others use sampling in different clusters so that they divide their population into different groups they call clusters and then use the clusters at random and from within those clusters or blocks they select their participants. Other researchers use weights to sample their participants. Say, you are studying within New Zealand and you know that in some areas, as in South Island, you will have less representative Maori population than in the North Island. You may want to weight your samples in a way that the people from Maori population are given greater weights than those from other ethnicity.

When you read research, pay close attention to how the individuals in the study were sampled. If the authors mention that they used simple random sampling, how did they actually do the sampling? From where did they select their participants? You can be assured that they used random selection if by reading their research methods, you know that they used a random numbers table to select their participants. Otherwise, you may need to think in ways their sample selection could have been biased.

We are getting into the issue of sampling bias. When you read research reports, also pay attention to from where did the authors or researchers obtain their participants? If you wanted to study the prevalence of attitudes of people towards consumption of fruits and vegetables, and you sampled individuals from affluent neighbourhoods and from malls from where people would obtain their supply of fruits and vegetables only, you might end up omitting or excluding people from the poorer neighbourhoods who may not consume much fruits and vegetables nor can they afford to consume or purchase them; you may be biased in your study only because you selected a biased set of people to provide your survey. These problems point to the fact that your sample is not representative enough. In occupational health studies, frequently, researchers survey factories and shop floors where they conduct cross-sectional surveys. If they conduct their surveys that way and only one time, they may end up oversampling those who were healthy to attend the workplace that day and miss the sicker workers. This bias is referred to as "healthy worker effect".

We wrote about the probability of occurrence of phenomena and finding the rare events. But how rare is rare? is there a borderline? What would be ways of thinking about some of these things? For those variables that are randomly distributed for example height or weight, you can use a normal distribution chart (Gaussian distribution chart) to map out the difference between what is expected and what is outside

12

of normal expectation. For example, let's say you are interested to find out the average height of year 7 school students in schools of Canterbury and for this purpose, you want to conduct a survey of a sample of 100 school children all over Christchurch. As you do so, you assume:

1. The school children in Christchurch are representative sample of all school children in Canterbury for that age group
2. If you have a sample of 100 school children, you will be able to estimate the average height of children for all of Canterbury
3. The height of the school children follow a normal distribution.

All of these are fair approximations and this is the good news: your measurement of the height of 100 school children in Christchurch will approximate the population average of all school children in Canterbury, but you also know that they will be off by a margin. This is the point of central limit theorem which states that the mean of the sample (m) will estimate the population mean μ ; the standard error about the mean (call it sd) will also estimate the standard deviation such that $\sigma = sd/\sqrt{(N)}$ where sd = standard error of the mean, and N is the sample size. Let's say after conducting the survey, you find that the average height was 150 cm and the standard error of the mean was 50 cm; then your estimated standard deviation would be $50/\sqrt{(100)} = 5$. How do we interpret this? The true population average would lie somewhere between 150 - 1.96 * 5 = 140.2 and 159.8 cm with the average of 150 cm. The magic number of 1.96 has come from the standard normal distribution (z distribution of a z score of 0.95. This means that 2.5% of the children will be either shorter than 140.2 centimetres or taller than 159.8 centimetres or 5% of the children will outside of this range. This concept that those who are outside of the 95% of the scores are "outliers" or statistically significantly different from the rest of the children is at the root of the thinking of a concept that the p-value of a distribution is 0.05 or lower. We will review this concept again in the following section on reasoning by abduction, but this is a concept well worth keeping in mind. The interval band of 140.2 and 159.8 in this case also tells us that if your survey were to be repeated 100 times, in 95 out of those 100 surveys, your average height of the children would be some figure in between 140.2 and 159.8 with the most likely figure around 150 cm. This concept is referred to as 95% confidence interval and we will review it in the section on statistical inference.

For now, we know that reasoning by inductive logic tells us:

- We obtain precise and accurate data about phenomena as "facts" using observational tools
- We then frame a "generalisable statement" about the pattern of our observation or on the basis of past experience and frame that statement in the form of a probability estimate or likelihood estimate

We can still use the framework of standard form where we use premises for facts or carefully observed valid and reliable observations and the conclusion in the form of generalisable statement. However, there are three things that we need to keep in mind:

1. First, our generalisable statement with probability statement on the basis of the observations must make sense. If they make sense and if it is OK to make that kind of a statement, then the probability of our generalised statement is high. A generalisable statement that is also likely to be highly probable is deemed as a strong argument.
2. Second, the value we assign to that probability estimate is called "inductive force" or "inductive strength". If both these conditions are met (that is the argument is both strong AND has high inductive force, then such as argument is referred to as a cogent argument.
3. Even though past experience may guide to frame our generalisable statement (albeit with certain probability estimate), by no means just because we experienced some events in the past, that implies the future will continue to be similar. We can have exceptions or events or observations that will not meet the generalisable patterns and this is the basis of refutation of the generalisability. David

13

Hume attributed this concept as "problem of induction" (see https://plato.stanford.edu/entries/induction-problem/)(phi). Nassim Nicholas Taleb (2007) expanded on Hume's observation and came up with a theory where he statistically modelled unusual or exceptional events that do not match with the expectations based on observations(Taleb, 2007). Such events are referred to as black swan events after the lore that there existed no black coloured swans in Europe and therefore it was unthinkable in the early 1700s that black coloured swans could exist. A Dutch expedition in Southern Hemisphere identified black coloured swans. Examples of black swan events in health sciences: in 2013, an outbreak of E.coli infections occurred in people who ate mussels cultured from the waters around Alaska. The event qualified as a black swan event in environmental health sciences as it was unthinkable at the time that E.coli outbreak from consuming mussels cultured in cold Alaskan water could have occurred; it was later identified that warm ocean currents circulated in the Alaskan coasts, rendering the E.coli infection possible(McLaughlin et al., 2005).

To recapitulate, reasoning based on inductive methods are based on first observations of events and phenomena and then generalising from these observed events, either in the form of predictions or general statements of truth. These generalised statements are in the form of proabilistic estimates of what is possible based on the observations. The extent to which the probabilities associated with the specific events and the generalisations are "feasible", we talk in terms of strength of the arguments; the actual probability of the generalisable "concepts" are referred to as inductive force and together they are referred to as cogent arguments. Cogent arguments in this sense are similar to sound arguments in deductive reasoning we reviewed earlier. That said, one observation to the contrary is enough to refute the general pattern of observation which are essentially based on conjecture. We will review this in details in the third aspect of reasoning: the reasoning by abduction.

Reasoning by abductive logic: explanation, theory, and hypotheses

Thus far, we have discussed that when we read research reports and review them, we pay attention to the individual arguments and we should attempt to evaluate their logical consistency and the force of these arguments. We test their arguments for their logical validity by constructing a standard form where we list the premises that contribute to the final conclusion (one argument, one final conclusion) and evaluate them. We also use reasoning by inductive logic to list premises in the form of systematic observation of phenomena and based on these phenomena, we arrive at a set of generalisations; these generalised statements are then expressed in the form of probabilistic arguments. While these allow for evaluation of individual arguments, these do not account for the reason or they cannot answer the question, "Why do we get to see the pattern that we get to see? Why or how such events occur?"

The answer to these questions come from a set of principles referred to as "abductive reasoning". In contrast with deductive reasoning that deal with the structure of the arguments and the validity of the arguments, and inductive reasoning that put a probabilistic perspective of a generalised truth on the patterns of occurrence of phenomena, abductive reasoning answers the question about what explains the occurrences. Abductive reasoning is based on three related concepts: explanation, theory building, and testing of hypotheses.

Explanation refers to a low level detailed approach to explain the phenomena that are observed. This follows directly from the probabilistic arguments in inductive reasoning. A more high level and abstraction of the explanation would be building a theory that would put a general concept on the explanations. A theory is a good theory if it can explain EVERY OBSERVATION by invoking some general principles. For a theory to be a successful theory, it MUST account for every observation related to an explanation. The theory would then must be tested for its robustness by searching for two things:

14

- The theory should be subjected to an investigation where an example or a condition must be found that does not meet the conditions of the theory or whatever is expected of the theory: that is, it must be an exception to the rule that the theory states. Thus, as the theory is based on a conjecture about the reality, we need an evidence that works to REFUTE the conjecture. All you need is JUST ONE counter example or evidence to reject the theory and search or frame a new theory. And/Or,
- We should search for a simpler theory than the existing one and compare the working of that theory with the existing one. A simpler theory would be one that would have fewer assumptions or fewer parameters. The fact that a simpler theory is a better explanation than a more complicated theory is attributed to William of Ockham (1287-1347). For an interesting account, see the Wikipedia entry here: https://en.wikipedia.org/wiki/Occam%27s_razor.

(note, in health care settings, Occam's Razor principle states that between two rival theories, one that has least assumptions and parameters is likely to be the simpler theory and therefore more accurate; in contrast to Occam's Razor, a rival hypothesis often used in Medicine for diagnostic tests is Hickam's dictum that states that most disease conditions have more than one causation, notably conditions such as the Saint's triad (hiatus hernia, gall bladder stone, diverticulosis) that all point to a related set of conditions, see more here in the Wikipedia entry here: https://en.wikipedia.org/wiki/Hickam%27s_dictum).

Essentially, you start with observations, and then frame a theory that should take into account ALL the facts and then search for other facts or situations that refute the theory. The way to do this is to use the theory to predict specific scenarios or set up hypotheses that are rivals of each other. The hypothesis that follows directly from the theory is referred to as alternative hypothesis, and the hypothesis that negates the alternative hypothesis and preserves the status quo is referred to as the null hypothesis. After stating the null and the alternative hypothesis, the researcher then collects data and tests whether the null hypothesis can be rejected so that the alternative hypothesis holds. The researcher does this by examining the probability estimates that his observations conforms to the conditions of the null hypothesis. If that probability is low, then the null hypothesis is rejected in favour of the alternative hypothesis. Convention states that the probability value at which one can reject the null hypothesis is 5% or lower. The probability estimate is also referred to as the p-value and is quoted in research documentation. Besides testing for the probability estimate at which the null hypothesis holds, the researcher also estimates the boundaries of the effect estimate using a 95% confidence interval. The 95% confidence interval states the boundary within which the effect estimate will lie, and states that if the studies were conducted a 100 times over, what would be the effect estimate in 95 out of those 100 iterations. If the effect estimate were to include the null value, then it would be reasonable not to reject the null hypothesis. Let's take a look at an example.

Example: In the late nineteen eighties, physicians observed patients from specific areas of the West bengal state of India with skin lesions that were characteristic of individuals who would be exposed to inorganic arsenic. At the time, the physicians developed a theory that the exposure to inorganic arsenic was through consumption of drinking water. The investigators set up a case control study design with people with and without skin lesions and assessed arsenic concentrations in their drinking water. They found that those with skin lesions had higher concentrations of inorganic arsenic(Haque et al., 2003).

Correlation, causation, inference, p-values, confidence interval

Let's recapitulate the story so far. Consider the following scenario.

P1: We have observed from geological sources that arsenic can be dissolved in groundwater

15

P2: If humans are exposed to inorganic arsenic, then they have a high chance of developing characteristic skin lesions

P3: Some humans were exposed to inorganic arsenic through occupational sources

P4: They developed skin lesions and other problems.

P5: Some people are exposed to arsenic in their drinking water

Implicit Assumption: Inorganic arsenic works in the same way if exposed through inhalation and through ingestion as in eating or through drinking

Intermediate conclusion: People who are exposed to arsenic in drinking water will have skin lesions

P6: Some people have developed skin lesions

C: While not absolute, there may be a probability that these people may have been exposed to inorganic arsenic through their drinking water

Up to this point, you get to see a play of inductive and deductive logic leading to arguments about exposure to inorganic arsenic and appearance of skin lesions.

Now we get into abductive reasoning and ask:

Why do some people exposed to arsenic in drinking water develop skin diseases (and some people do not develop skin disease)?

With a question like this, there are several lines of explanations and theories that are possible. Some (not an exhaustive list):

Explanation 1: There is a threshold limit above which arsenic will cause skin disease and below which arsenic exposure will not lead to skin disease (call it dose-response theory)

Explanation 2: Arsenic may act differently in children and adults: so while adults may get the skin lesions, children may not get it (arsenic age theory)

Explanation 3: Arsenic may take a long time to act and so for some people the lesions may show up and for others it may not have shown up yet (arsenic time theory)

Explanation 4: Some people may metabolise or remove arsenic faster than others and therefore for these people arsenic may not cause skin lesions (arsenic metabolic theory)

Explanation 5: Some people may metabolise arsenic faster and it is the metabolic by products rather than inorganic arsenic as such that can lead to diseases (arsenic metabolic byprouct theory)

. . . and so on.

A few things to note here:

- Each of these theories must be able to explain every observation noted so far (in our case why some people develop skin diseases while others do not and that arsenic exposed through drinking water will cause skin diseases)
- You can name the theories any way you like, and it is convenient to start with an explanation first
- For the same phenomenon, you can have multiple theories and theories can be directly opposite of each other. For example, if you read the last two theories, you see that they are directly opposite of each other the way they are framed.

Once you have the theories, then you should start putting together hypotheses. Think carefully about the hypotheses that you will put up. The first hypothesis, termed the alternative hypothesis will conform to

16

the conditions of your theory and will be stated as such. Let us start with the first explanation or the first theory the "threshold theory" and see how we can build up a pair of hypotheses from this theory by way of an illustation.

So you note that this theory tell us that the reason some people develop skin lesions and some other people do not develop skin lesions even though all of them are exposed to drinking water that contains arsenic (albeit in various concentrations and people differ in their daily water intake) is that, arsenic may act on the basis of a threshold value. This indicates or implies that if someone were to drink or consume arsenic at low concentrations, that person would not develop skin lesions. Which lends support to develop a dose response study with arsenic exposure, that is, people in lower dose of arsenic in the body will not develop skin diseases while people with higher dose of arsenic in the body will develop skin diseases. Remember also that for every theory is as good as how we refute it. In other words, you cannot "prove" a theory, you can only fail to disprove it. The way to disprove a theory is to either find a counter example or propose a hypothesis that will render it false. Karl Popper, the Austrian philosopher of science, who spent some time at hte University of Canterbury referred this as "falsification of hypothesis" (see his essay at http://stephenjaygould.org/ctrl/popper_falsification.html. This is also referred to as Karl Popper's theories of conjecture and refutation (Popper, 2014). Anyway, if you get the idea, let's put the alternative and null hypothesis together:

Alternative hypothesis (H1): People exposed to higher levels of inorganic arsenic in drinking water will have a higher likelihood of skin lesions, or alternatively, compared with people without skin lesions, those people who have skin lesions will have higher levels of inorganic arsenic in their drinking water

Null hypothesis (H0): People with and without skin lesions will have similar levels of arsenic in their drinking water, or alternatively, the risk of skin lesions will remain same for those with high levels of arsenic in drinking water and those with relatively lower levels of arsenic in drinking water.

This is it. Note the sequence of how it all works together. You have made careful observations or studies of phenomena. You use reasonings of abduction or abductive reasoning to explain the phenomena, and you set up theories (well, explanations first and then theories). One theory, one set of hypotheses. Your hypotheses are tied to the theories you have about phenomena you want to explain. You may be tempted to put up several hypotheses to test but remember that your hypotheses must be derived from the theory you want to test. Rival theories will have rival hypotheses or you can use the same pair of hypotheses in ways to test rival theories. Let us stick to the simpler notion that we will have one pair of hypotheses for one theory.

Now based on the hypothesis you derive from your theory, you will set out to collect data. Your data collection can be on the basis of surveys that you will conduct yourself, or administer surveys in many different ways, or you will conduct experimental studies, or other means, for instance, compile other studies to conduct a meta analysis. After you will have collected data, you will examine whether your findings can be explained on the basis of your null hypothesis.

But how will you know that your null hypothesis is to be rejected or whether you fail to reject the null hypothesis? You set it up before the study begins. Consider the following table:

Null Hypothesis (True or False)

Test Results	Null True	Null False
Reject Null	Type I Error	Correct
Fail to reject Null	Correct	Type II Error

This table shows comparison between what will the test results show (that is the study results here are referred to as "test results") with respect to the status of the null hypothesis. If the null hypothesis is true but the study results suggest that the null hypothesis should be rejected, then we have committed an error. This error is "false positive error", that is, we have obtained a positive result in favour of our alternative hypothesis and it may lead us to believe that our theory is correct, but in fact that is not the case. Hence we label it as Type I error. This is also referred to as α error; we must specify the error rate before the study begins. Depending on how confident we are on our hypothesis, we can set it to any value. By convention, it is set to 5% error; by setting it so, we state that if we were to conduct our study over and over again, in 100 iterations, we would at most commit 5 times that our findings will not match the null hypothesis truth, that is if the null hypothesis truly were to be rejected, in 95 times out of 100 iterations, we would be able to reject it.

Compare this with another error. This time we commit a false negative error and we term this as β error; here, we will collect data and analyse and then realise that we cannot reject the null hypothesis but the null hypothesis should actually be rejected as it is false. We have committed an error and we can set the error rate at something like 20%, so we state that if we were to conduct 100 iterations of this study, we may be wrong this way about 20 out of those 100 times, and we would call that we fail to reject the null when the null is true 80 out of 100 times. That also means that if the null were to be truly false, we would correctly reject the null 80% of the time. This is not the same as that if we were to reject the null, we would wrongly reject the null only 5% of the time. So this figure that we'd correctly reject the null WHERE the null is false is our power of the study. The alpha error, the power of the study we aim for, and the effect size that we would like to study together are factors for us to estimate the required sample size for any study we'd like to conduct. But note that we'd have to call these figures before the study begins.

Then we conduct the study and we obtain some figures to compare with the null. This depends on what effect size we would like to compare for the study we conduct. If we want to settle for odds ratio or relative risk estimate on the association between two variables (exposure variable and outcome variable), then under conditions of the null, the effect size would be 1.0. On the other hand, if we wanted to study the difference between two measures, then under conditions of the null, the effect size would be 0. We also assume as we conduct our study that this study is one of the myriad possible iterations, so the effect size we obtain here should follow a normal distribution pattern and therefore we now estimate the probability that the effect size we have obtained falls within the expected distribution of the null estimate. Note that the null's point estimate is 1.0 (for ratio type measure) or 0.0 (for difference type measures) but taking that as a point estimate, the null hypothesis could predict that the band around the null value could take any other value, it's just that their point estimate will be 1.0 or 0.0. Does our point estimate and the band we construct around our point estimate fall within that boundary? Or is it far out? If far out, how far out?

For example, let's say we conduct a study on arsenic exposure through drinking water at various levels and the risk of skin diseases. Further, for illustrations, let's say we would like to compare the risk of skin lesions for people who were exposed to inorganic arsenic at 50 ug/L or less and those who were exposed to levels like 200 ug/L or more. Then we fix that those who were exposed to the lowest level of arsenic exposure (that is 50 ug/L or less) as the reference category and say that at that level, we will consider the risk to conform to null value so the odds ratio will be fixed at 1.0 (that is no risk). Then, based on our sample, we compare people who were exposed to arsenic at 200 ug/L and those who were exposed to 50 ug/L or less. The effect measure is Odds Ratio and let's say we get a point estimate for our sample to be 2.5; we interpret

that compared with those who did not have skin lesions, those who had skin lesions were 2.5 times likely to be in the arsenic exposure group of 200 ug/L or more. But this is just part of the story as we only have this one sample of people on whom we have based this. If we were to conduct this study a 100 times over, what would the distribution of the effect measure look like? So we calculate that as well, and let's say we find that range to lie between 1.5 through 4.5 for 95% of the 100 iterations. Based on these figures we state that the best estimate of the population odds ratio for the association between arsenic exposure at highest level (that is 200 ug/L or over) versus arsenic exposure at lowest level for skin diseases is 2.5 with 95% confidence interval band of 1.5 - 4.5; If there are other values, they will lie at the 5% extremes. We then go ahead to state that there is a statistically significant risk of skin lesions due to high arsenic exposure. We can also estimate the probability that our point estimate would be compatible with the null hypothesis? Let's say we get a figure of around 2% or there is a 2% chance that such a figure of effect size could be possible under conditions of the null. We then know that we have successfully rejected the null hypothesis at 5% level and we express this probability estimate as p = 0.02. This is the "p-value". So you can see that the p-value does not tell us any more than the story that there is a probability value we can put to the findings if the null were to be true but we do not know anything about the distribution of the effect size or any other value to be meaningful. But one could deem that the study is statistically significant. We will leave at that, and I encourage you to read more about the issues around p-values. A particularly useful paper to review about p-values is Andrew Gelman's commentary (Gelman, 2013).

You can see that p-values and 95% confidence interval estimates solve the issue whether an association we observe could be spurious or could have arisen because of chance factor alone; you now know that if the association were to be statistically significant, you could argue that a statistically significant association would settle or rule out the play of chance. But that just settles one aspect when it comes to health related research: there are two other issues that we need to settle before we can say that the association we observe between an exposure or an intervention and an outcome (health outcome) is one of true association. These are biases that are present in a study that can make the study invalid, and confounding variables that were not adjusted for: again rendering the association open to suspicion. But before we delve into those, let's take a look at issues around spurious correlations.

Spurious correlations & ecological fallacy

Spurious correlations are those where you have two variables that seem to be associated with each other and the associations cannot be just due to chance, yet it makes no sense to have these sort of associations. I recommend you review a website (and a book that is linked to the website) by Tyler Vigen (Vigen, 2015) here:

http://www.tylervigen.com/spurious-correlations

Take a look at the first chart from the book:

You can see that the two trends, that is, suicides and US spending on science and technology over time have very similar trends. If you were to plot these data together (say between 1999 and 2009), you would even see a linear association on regression. Does that mean that increased spending lead to suicides or if you reduce spending on science and technology, suicide rates would come down, :-)? Can you think what must be going on here?
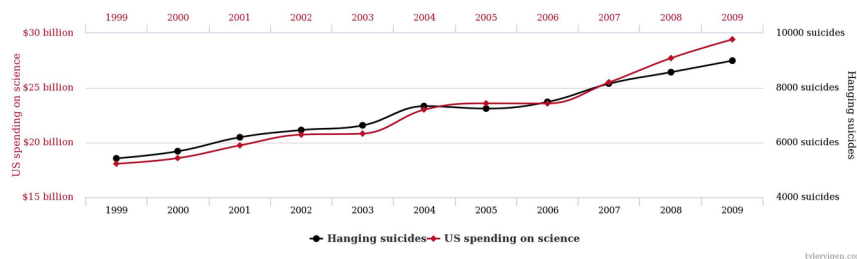
19

Figure 1: Association between suicide and US spending on science/space/technology (from Tyler Vigen's Spurious correlation website, see text for link)

So just because X and Y are correlated with each other does not necessarily imply that X is a cause of Y, nor Y is a cause of X. Had that been the case, you would be able to argue for reduction in the US spending towards science and technology to reduce suicide rates. But why cannot we argue when we see this sort of correlations or what makes these correlations unfit for any use? Well, if you see correlations such as this, you need to ask yourself, why do we get to see these correlations? Is it possible that there is as yet a third variable that can account for both the trends and once we "control" for that third variable that is associated both X and Y, this apparent "association" will disappear? What could be that variable in this case? What is your theory? Could it be that societal complexity and technological advancement that has resulted from increased funding have resulted in dissatisfaction among certain section of people who decide to commit suicide and the pace at which this happens matches the pace of sci/tech funding? Can we study them? How can we study them? What do we need to know?

Here's another issue. Consider the following figure. This figure has come from a study conducted by Xie et al (2014) where they studied morbidity and mortality from heart disease (ischaemic heart disease) in Beijing, China (Xie et al., 2014). They studied aggregated measurements of air quality in China and aggregated number of people who died or were admitted to the hospital with ischaemic heart disease in that city in 2014.

This figure (Figure 2) suggests that with increasing concentration of fine particles in air, the corresponding morbidity (illness) due to ischaemic heart disease also increases in Beijing. The numbers for this graph were generated by using an ecological study where aggregated measures of fine particulate matter and ischaemic heart disease related hospitalisations were analysed. Based on this data, let's say someone already has ischaemic heart disease. Is he more or less likely to get hospitalised due ti IHD on a day when the particulate matters are very high? We cannot say with certainty, because studies where aggregated data are used to make predictions one cannot extrapolate the findings to indiviuals: this is known as ecological fallacy - that is, you cannot extrapolate results from aggregated data to individual cases.

So spurious correlations and ecological fallacy each point to the fact that not all correlations can be causal and for causal linkages, the associations must be valid for individual cases. A valid association must account for three situations:

- The association that we observe must rule out the play of chance. – We have seen how we can start with theories and set up hypotheses based on the theory and then use the hypotheses to identify the
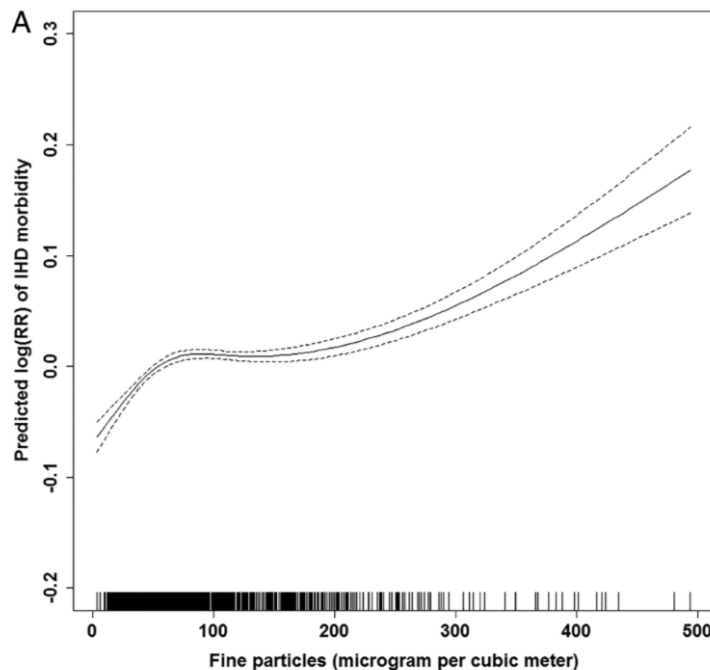
20

Figure 2: Relationship between fine particulate matter in air and risk of illness from ischaemic heart disease (IHD). Source: Xie et.al (2014) study

alpha and beta errors so that we can deal with adequate sample size and power estimations that will allow us to rule out the play of chance.

- The association must eliminate all possible biases. – Biases refer to systematic measurement errors between comparable groups that are studied in an association study. For example, let's say we are planning to study the association between cigarette smoking and lung cancer; for this purpose we have decided to study the prevalence of smoking among those people with lung cancer and those who do not have lung cancer. Further, we have decided to measure their extent of smoking by asking them using a questionnaire. In a setting like this, we can have many different biases. Bias will depend from where we have sourced our participants. If we have selected lung cancer patients from cancer wards and non-cancer patients from say a gymnasium where relatively fit, younger, and health-conscious people attend, we would have introduced a selection bias. This would be a bias where by selecting people in different way who will then be compared the investigator has already introduced a bias in the study. To avoid selection bias, we need to make sure that the people we select for our study in the comparison groups must be as similar as possible. One rule of thumb is that, in a case control study, if the control were to develop the disease of interest, then he would be a case, they should be so similar. So, for the smoking-lung cancer study, we could have selected our controls from another ward in the same hospital (perhaps another inpatient ward) and people with similar age and social profiles. As we have not used an objective measurement for measuring their exposure, this would still introduce other forms of biases. For example, as we are surveying people with and without cancer, the way they may relate their extent of smoking may be different. Perhaps the smokers with cancer are may remember their smoking histories better than non-cancer patients. This form of bias is referred to as "response bias". Biases must be eliminated at the stage of planning the study, as following data collection, the biases may have already occurred, and one can do little about elimination of biases after data collection.
- The association, to be valid, must control for confounding variables. – The term confounding variable refers to the situation where a third variable exists between the exposure and the outcome (or

21

between the intervention and the outcome) such that (1) this variable is related BOTH to the exposure/intervention AND the outcome variable, and (2) while this variable is related BOTH to the exposure and the outcome, it does not come in any way where a causal path can be constructed between them. For example, imagine we are studying the association between smoking and risk of heart disease. Review by Karen Matthews and colleagues suggest that premenopausal women have lower risk of heart disease compared to men (Matthews et al., 2009); and as Okene (1993) suggests, prevalence of smoking among women is lower than that among men (Ockene, 1993). But gender, even though related to both the exposure and outcome, does not come in the causal pathway that may connect smoking with heart disease. Therefore, gender is a confounding variable and must be treated as a confounding variable in any research connecting smoking with heart disease. You can control for confounding variables in the planning stage of your research or at the stage of data collection, and indeed during data analysis. In the planning stage or before collection of data, depending on the type of study, you can (1) randomly allocate your participants into the intervention and control groups (if intervention research such as a randomised controlled trial), or (2) you can restrict your participants to one specific level of the confounding variable (for example if you were to study smoking and heart disease association, you could only work with men); alternatively, you could match the participants on the basis of the confounding variable (you could maintain a 1:1 ratio of including women for both comparison groups in the smoking study); at the stage of data analysis following the study completion, you could stratify the participants according to the categories of the confounding variable and then pool them in the final analysis, or (2) you can conduct multivariable analysis where the confounding variables are entered as variables in the statistical data analysis.

Thus, only after you have ruled out the play of chance by framing the hypotheses, conducting a sample size estimation and power analysis, after eliminating all biases, and after controlling for the effects of confounding variables, if you find that an association persists that is both substantively (that is theoretically acceptable) and statistically significant, then you can claim that a true and independent association exists between the exposure/intervention and the outcome variable. But it still does not answer the question whether such an association is causal or non-causal.

Not all validated associations are causal in nature: as Rothman and Greenland (2005) have argued, most causal models are multifactorial (Rothman and Greenland, 2005b). What this means is that, a disease or an outcome has more than one cause and causes interact with each other. The fractions that we attribute to a specific factor as a cause for an outcome, when we add them together, can be lower than 100% or may be higher than 100%; this is because different causes interact with each other. Some causes are necessary causes such that, in their absence, a disease will not occur. For example, for tuberculosis to occur infection with bacillus tuberculosis is necessary; without the bacilli, signs and symptoms of tuberculosis will not manifest; even then, many people have been infected in the past with tb bacilli but they do not manifest the disease unless they have compromised immunity or some other conditions that would enable manifestation of tuberculosis or other risk factors; for example review Faustini's systematic review on the risk factors for tuberculosis (Faustini et al., 2006). This suggests that a causal variable can be necessary but rarely sufficient by itself to cause disease. On the other hand, when the necessary causes "team up" with other variables that contribute to the emergence of the disease, a "sufficient causal model" emerges. You can create several sufficient causal models of disease conditions, referred to as "causal pie" models (Vineis and Kriebel, 2006).

So in the context of health sciences, what factor may be considered as a cause for an outcome is not often clear and making that call is fraught with subjective limitations. In 1965, at a conference of occupational hygienists in London, Sir Austin Bradford Hill (1965) proposed a set of nine criteria (Bradford Hill, 1965). Although Sir Hill did not mean these as "criteria" but considerations, these have come to be known as "Hill's Criteria". These considerations help us to assess the nature of an association between an exposure and a disease outcome. These include:

1. Strength of Association. – If an exposure or an intervention is associated with an outcome, then how

strong is that association. By strength of association, we mean if we review the odds ratio, or if consider the relative risk (risk of the outcome in the exposed population or those who have received intervention divided by risk of outcome among those who have not received intervention or those who were not exposed), or absolute risk (that is risk in the exposed minus the risk of the outcome in the non-exposed), what is the magnitude? Hill discussed in the context of smoking and lung cancer or heart disease risks and those risks were in the magnitudes of 10 or higher; but the purpose we should look into strength as a measure to judge whether the association is one of cause and effect is that, the stronger an exposure will be with an outcome, the tighter will be the link: you will need to think of another highly prevalent exposure or a stronger factor to have that strong an association. Also, a strong association would indicate that a substantial part of the outcome can be attributed by the exposure.

2. Consistency of association. – If we study the linkage in different populations and different circumstances, do we get similar results? This would suggest that if we were to see very similar pattern of associations in different population groups, that consistency would suggest that this association is not spurious but substantive.

3. Specificity of association. – What this means is, if there are specific contexts where the associations become prominent, there could be a cause and effect linkage. For example, consider a workplace where some workers are exposed to high concentration of environmental tobacco smoke (ETS) and in the same workplace, in other areas are relatively free of the ETS. If we get to see after studying these groups of employees that those who worked in the ETS areas were more likely to suffer from chronic lung disease, then this raises a suspicion about ETS being a causal factor for chronic lung disease.

4. Temporality of association. – If X is a cause of Y, then it makes sense to think that in the chain of causation, X has to happen earlier in time than Y. If that is not the case, then it is hard to justify X as a causal variable. Hill thought that this was like "putting the horse before the cart", and Ken Rothman (2005) contend that this is the strongest clause for cause and effect estimation (Rothman and Greenland, 2005a).

5. Biological Gradient of the association. – We know biological gradient as dose-response assessment. This means that as the dose of the exposure increases, or as the intensity or frequency of exposure increases, so will be a corresponding change in the outcome. We can argue that this may not hold true for all cases; particularly for exposures that have ceiling effect (that is where, the exposure reaches a high point), you may not experience a corresponding rise in the outcome as well. Read the work by Philips et.al, (2006) for a comrehensive review on the causal criteria (Phillips and Goodman, 2006).

6. Plausibility as a criterion. – In his lecture Hill mentioned that it would be "helpful" if the cause would biologically account for the outcome. As we know from cases, this may not always be the case. Quite often, we may not know in advance the biological basis of an association nevertheless there can be a causal linkage. Think of John Snow's investigation of the London Cholera outbreak. While he blamed the water supply of a specific company, at the time of his investigation, no one knew about cholerae vibrio and their roles in the London outbreak. See for example, vandenBroucke's account of John Snow's cholera outbreak investigation (Vandenbroucke et al., 1991)

7. Coherence as a causal criterion. – This basically means that are there other instances where similar associations are found? This follows Mill's canons of induction. Read Ducheyne's discussion on the matter (Ducheyne, 2008), and as we have seen, if we see that there are similar associations albeit in other domains, then that confirms our ideas that we may be experiencing the action of a causal agent. For example, having known that inorganic arsenic was a known factor for bladder cancer, Smith et.al. (1998) proposed that exposure to inorganic arsenic in drinking water could also be a risk factor for lung cancer (Smith et al., 1998)

8. Experiment as a causal criterion. – If we know that X is a cause of Y, can we set up an experiment where we can introduce or control X and we will see a corresponding reduction or change in the status of Y? Experimental validation is not always possible in case of health sciences, and indeed, in clinical context, we can think of randomised trials where one condition can be deliberately controlled to test whether the association is one of cause and effect

9. Analogy as cause. – Are the exposure and outcomes analogous to some other organisms, other species that produce similar situations with respect to disease causation?

None of these are hard and fast rules for setting up a cause and effect assessment. Each of these nine criteria, except perhaps for the temporality, is open to refutation and criticism. Nevertheless, these provide us with a guidance as to assess whether some factor is associated with something else in the manner of cause and effect. It is judgemental, and qualitative but these provide us ways to think about the association. You will indeed find that in many articles and presentations, authors/researchers refer to these criteria when they assess the cause and effect nature of the associations. Indeed in 'reasoning by abduction' where our goal is to assess "why we get to see the patterns", explanatory models, theories, hypotheses, and discussions of cause and effect associations that differentiate from other types of associations are beneficial for the assessment of the literature.

So this brings us to the beginning of the final topic on our journey to research methods in health: that of what study designs shall we adopt for our study questions. Also, when we assess studies and when we propose our own studies, what study designs might be appropriate and feasible. Let's take a look at the different study designs we can use and their pros, cons, and indications. We will discuss basic features of each study design, where they can be used, what are their advantages and what are their pitfalls that we must be aware of as we either assess in studies where they appear or decide to use them in the context of our own research.

Study designs in Health

It helps to think in terms of either intervention research or observational research. Alvan Feinstein (1997) has argued that this dichotomy ("observational" versus "interventional") is useless as constructs as everything is observational; for instance, do we not conduct observations even in the context of experiments or randomised trials? But here we will therefore use this only in the spirit of separating the notions of "what we do", not so much in the spirit of a container of two different kinds of studies (Feinstein and Horwitz, 1997).

Study designs that are neither strictly interventional studies nor observational but something else

In this class of studies we put systematic reviews, meta analyses, secondary data analyses, and ecological study designs such as time series or spatial studies. These are as follows:

Systematic reviews. – These refer to a process of (1) framing an answerable question; (2) constructing a search algorithm to identify relevant studies, (3) selecting and rejecting studies based on criteria that the authors/researchers set up before the beginning of the study; (4) abstracting information from individual studies using a plan on a spreadsheet or tables; (5) assessing the methodological qualities of the studies, (6) summarising the results of the individual studies to a series of answers to the questions that the authors set out to study. Systematic reviews are best used when you are starting out or when some investigations are already done by others and you want to summarise "what is out there". It is always a good idea to systematically summarise the key messages. It is a starting point for most research. However, the quality of an SR is as good as the constituent studies; when you conduct SRs, you are also limited by publication bias where you are biased by only those studies that are published and have positive results; you will need to identify other studies and fugitive literature (studies that are not published) for robust systematic reviews. See Campbell Collaboration website to learn more about systematic reviews here (col, a): https://campbellcollaboration.org/

Meta analyses. – These class of studies are virtually identical with systematic reviews and follow the steps till step 5; then the summary of results in meta analyses follow statistical procedures where the studies are first assessed whether they are "homogeneous or heterogeneous" – these terms imply whether the studies on the basis of the population, or the results belong as if they are similar in a way. Accordingly a number of different ways in which the summarised numerical results are presented. The results of individual studies are combined statistically to arrive at a summary estimate. Like systematic reviews, these studies are open to publication bias. Meta analyses are conducted when you have data on randomised controlled trials on interventions for specific health outcomes. See for example Cochrane Collaboration (URL: http://www.cochrane.org/) to learn about how meta analyses are used (col, b).

Secondary analysis of data. – Large data repositories are increasingly becoming available where you can access data sets other researchers have collected or governments have collected and they have made them available for researchers all over the world. You can use tools of biostatistics and data analysis to mine those data bases. Kaggle for example, provides data sets that you can use to analyse (see URL: https://www.kaggle.com/); in the early stages of your thinking about a topic, it is a good idea to delve into secondary data sources and analyse them. It is also a good idea to test theories and hypotheses when you cannot collect your own data or find data collection very expensive. Secondary analyses of data are excellent sources for developing theories and validating theories. **Ecological studies**. – These are a subset of secondary analyses. Government data bases allow you to analyse associations between various factors: for example, you can study area level deprivation and health outcomes in New Zealand and this will allow you to understand and test theories on association between poverty and health issues for instance. The limitation of ecological studies and secondary data analyses that are based on aggregates of data are ecological fallacy where you cannot extrapolate data on the basis of aggregates to the individual cases.

Primary interventional study designs in health care

Randomised controlled trials. – These are primary studies in the sense that you collect data from individual participants; these are interventions where you randomly allocate participants into intervention arm and control arm. The studies can be blinded so that you may not know which participant received the intervention and which participant received the control condition (single blind), or you can be blinded as well as the participant (double blind trials); randomisation and blinding ensure that you have taken care of confounding and bias inherent in the study design. The effect estimate is in the form of absolute risk or risk reduction (risk of outcome among those who received the intervention - risk of outcome among those who received the control condition), or in the form of relative risk estimates. RCTs are conducted in clinical contexts or even in population level or community level interventions to test the efficacy of interventions under "controlled conditions". These studies are as close as possible to the experiments on humans. The studies are limited by the fact that these are expensive studies and the results are not generalisable to all members of the public: the results of RCTs apply to individuals whose profiles match closely the profiles of those who were included in these studies. This is why RCTs and intervention trials have limited "external validity" while they have excellent internal validity.

Study designs that are mainly observational in nature

Prospective cohort studies. – These are classes of observational epidemiological studies where the investigator assigns the participants of the study into groups where the participants are exposed or non-exposed to an exposure of interest. The investigators then follow up the participants over a period of time till the participants develop the outcome of interest. The rates at which those who are exposed and those who are non-exposed develop the outcome of interest are compared and these comparisons form the effect estimate. The effect estimates are either absolute difference in the rates at which the outcomes occur (ARR or absolute risk reduction or Attributable Risk), or Relative Risks (the ratio of the rate of occurrence of the outcome in the exposed versus in the non-exposed), or Hazard Ratios (hazards are defined as instantaneous risks of

25

occurrence of health events and therefore hazard ratios are the ratios of the hazard of the outcomes among the exposed and the non-exposed. In a prospective cohort study, the participants are identified and followed up in present time and the outcomes may occur at a later date. For example: the researcher is interested to study the incidence of acute respiratory tract infection in children who are born in regions that are close to gold mining areas or away from gold mining regions in a country (the idea being to test if being exposed to a mining environment is associated with ARI in infants). So they followed children (exposed cohort) who were born and lived for the first year of their lives in gold mining areas and those children who were born and lived in the first year of their lives (as infants) in non-gold mining areas and noted the relative incidence of acute respiratory tract infections in both cohorts. Prospective cohort studies are useful for studying multiple exposures and commonly occurring illnesses or health conditions; they are useful study designs to assess causal linkages between exposure and outcomes; however, they are time consuming and expensive and not suited for studying rare disease occurrences.

Retrospective cohort studies. – Retrospective cohort studies share the similarity with the prospective cohort studies that in both types of study designs the investigators follow up individuals with defined exposure status (these individuals are referred to as "cohorts", so there are exposed and non-exposed cohorts), but in case of retrospective cohort studies, the exposure and the outcome have already occurred in a past time frame. The investigators know and have identified on the basis of pre-existing records as to who are exposed cohorts and who are non-exposed cohorts and can recreate the emergence of health effects. Retrospective cohort studies are used in industrial and workplace setting and in occupational epidemiological settings to study the emergence of health outcomes among individuals exposed to different levels of a toxin in the workplace. Like prospective epidemiological studies, this is a good study design for causal inference; however, as in the case of prospective cohort studies, these studies can be expensive and you can only study common outcomes (so, for instance, occupational cancers are not good candidate research topics for retrospective cohort studies).

Case control study design. – Case control study designs are used to study the likelihood of exposure for individuals with known states of health outcomes. In case control studies, researchers first identify identify individuals with a disease condition ("cases") and similar individuals selected on the basis of potential confounding variables, who are free of the disease under investigation ("controls"). For both cases and controls, the investigators then study the likelihood of their past exposure. The likelihood of past exposure is referred to as Odds of exposure; the odds of exposure for cases are compared with the odds of exposure for controls, and hence the effect estimate is odds ratio of exposure for cases and controls. Case control studies are used to study rare diseases (where you already know the status of the individuals with the disease and without the disease), and they can be used to study multiple exposures for a simple set of outcomes. Case control study designs are perhaps the most widely used epidemiological study designs where the investigators can identify individuals with and without disease states and then can identify the likelihood of exposure for each group. Case control studies are used for Cancer Epidemiological studies as cancers are relatively rare disease conditions (rare diseases are assumed to be diseases that occur in 1 per 10, 000 individuals or less frequency); case control study designs are open to response bias and it is difficult to assess causal linkages using only case control study designs. For example, if you set up a case control study design to assess the association between smoking and heart disease, and find that those individuals with heart disease have lower rates of smoking, you will not be able to know for sure whether the lower rates of smoking among individuals with heart disease occurred after they found out that they had heart disease or not.

Cross sectional surveys. – Cross sectional surveys are also referred to as prevalence studies. In cross sectional surveys, investigators study at a single given point in time or a single given time period a group of individuals (or a sample of individuals) on disease outcomes and their exposure status. The study design is suited to estimate the prevalence of a health state. For example, you can use a cross sectional survey on a group of individuals to estimate the prevalence of diabetes in Christchurch or Canterbury or even whole of New Zealand. The cross-sectional studies can also be used to model prevalence odds ratios for those individuals with a defined disease state and those without the disease. However, cross sectional surveys are limited by the responses or measurement errors on parts of the investigators and respondents or participants.

26

Cross-sectional surveys are also poor study designs if you want to assess cause and effect associations between exposure and outcomes because you are going to obtain the exposure and outcome information from the study participants at the same time. As temporality is a major determinant of causation, therefore you cannot be certain if the exposure definitely preceded the outcome. Other than that, Cross sectional studies are relatively inexpensive in terms of the amount of resources they consume and the time you will take to set up such a study.

Case series. – The previous study designs can be used for studying cause and effect relationships between an exposure and an outcome (or an intervention and an outcome); this is because each of these study designs include a valid comparison group in one way or another: controls who receive standard interventions or placebos in RCTs, non-exposed groups or cohorts in cohort studies, controls in case control study design, and indeed, among individuals who are either interviewed or measured for specific health events or effects and exposures in cross sectional study designs. Case series, unlike these study designs, is about description of individuals with specific exposures and/or health effects. In case series, a number of individuals are described with respect to their disease conditions either at one point in time or over time periods. Surveillance conducted by governments and health departments to study emergence of disease conditions (both infectious disease surveillance and non-infectious and chronic disease surveillance) are examples of case series in action for public health. Case series is inexpensive and the case series study design helps to generate hypotheses or form the beginning of abductive reasoning that leads to answers for further questions about why we observe the pattern we observe using a case series as study designs. Case series are not good study designs if our aim is to find an answer "why" or "how" of a disease or disease-exposure association. Otherwise case series are relatively inexpensive and the least resource intensive of all the study designs we have covered so far.

So abductive reasoning is about answering an explanation for a series of events we observe: the explanation is first offered on the basis of intuition or conjecture. Based on the explanation, the researcher frames a theory and based on the theory, the researcher then sets up rival hypotheses. The hypothesis that immediately follows the theory is termed as "alternative hypothesis"; the hypothesis that challenges the alternative hypothesis and is related to a state of status quo, is termed as "null hypothesis". The rules of framing a theory is that all relevant observations must be taken into account and theory must be able to explain EVERY fact that are observed. More than one theories are put up for examination; the theory that fits all observations and is also the simplest or involves least number of parameters, being simple, conforms to the Ockham's razor principle and wins; but this is not hard and fast. As a researcher, another goal is to find out a fact or evidence that refutes the theory and comes up with an observation that is in contradiction to the predictions of this theory: this principle is therefore referred to as "conjecture and refutation". Using abductive logic, a researcher will claim that he or she cannot prove a theory but only fail to disprove the theory.

Now that we have covered the three classes of reasonings (deductive, inductive and abductive), let us put everything we have learned so far into practical use. We will do this (1) first by outlining ways in which we can ask questions and abstract information from papers we read, and (2) how we can use this approach to develop research proposals.

First: How to read and assess research

The first thing you should do is to think of a research idea. What is a research idea? We define here or describe here a resesarch idea as one that needs an explanation. In order to do so, we need to take stock of existing knowledge base and we should have a way to address these first. Hence I recommend the following steps.

Step 1. Run a search and do forward and backward citation tracing, reach a concept saturation point, and review the articles

Citation tracking refers to the process of identifying which citations are linked to which ones and thus you can find the most influential article on a topic. Raul Pacheco Vega tart with identifying a set of citations that are reviews if the topic is well studied, otherwise use citation tracking for a set of studies or published reports going no farther than say the last five years, for more details, see http://www.raulpacheco.org/2018/02/forward-citation-tracing-and-backwards-citation-tracing-in-literature-reviews/. Example:

We would like to study the association between heat waves and morbidity and mortality. We ran a search on Google Scholar (URL: http://scholar.google.com) without restriction on any date. Note the following two figures:
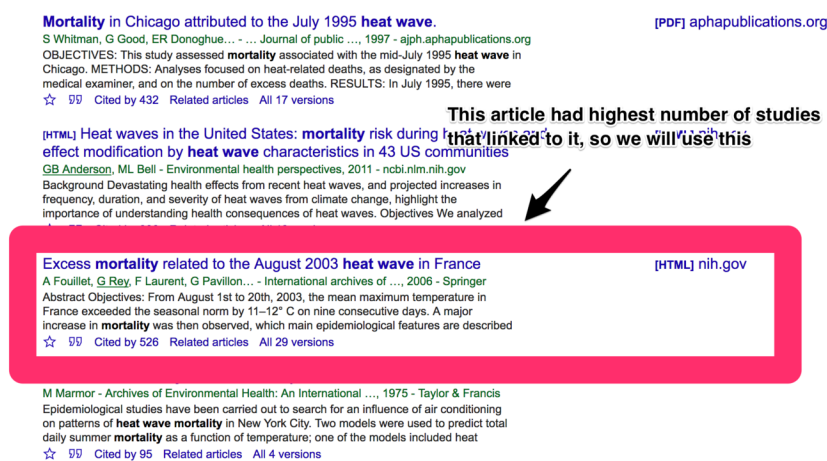

Figure 3: When we searched Google Scholar with mortality and heat wave, this article came up with maximum number of articles cited it. This is referred to as forward citation tracing. This suggests that this article has been widely read and cited and may contain important information worth reading. Now note what happens if you click this link to identify which articles refer to it.

So, this is forward citation tracing that helps you to identify which articles are very influential and which are highly cited in the literature and therefore likely to contain information that scholars have found useful to cite or read. A good idea not to overwhelm you with information overload would be that, you select the last few years of publications and continue with forward citation link search to identify at most four or five articles that will likely to contain the information that you'd like to focus. Your next thing would be to conduct a backward search of the citations. For this, you review an article (either the seminal article that set you to the path of search or an article that looks highly cited and see which other articles did this article cite. Then read a few of the articles in the reference list of that article and gradually using this process, you start developing a sense of which concepts or conclusions, or arguments are worth chasing and pointing to, which we will turn next. You may want to review Raul Pacheco Vega's directions he has posted in his well read blog here: http://www.raulpacheco.org/2016/06/how-to-do-a-literature-review-citation-tracing-concept-saturation-and-results-mind-mapping/

The following figure illustrates the principle of backward citation tracing:

This way you go through forward and backward citation tracing a number of times till you get to see a set of articles and authors being cited over and over again. Pacheco-Vega (2018) calls it concept saturation where

28

This is FORWARD citation tracing as you see which articles have cited the original article since its publication and this, for instance, leads you to a very widely cited article that is also a review - this may give you pointers to the questions you ask: how does temperature affect health? or Why do people die from heat? Or what explains public health impacts of heat waves?
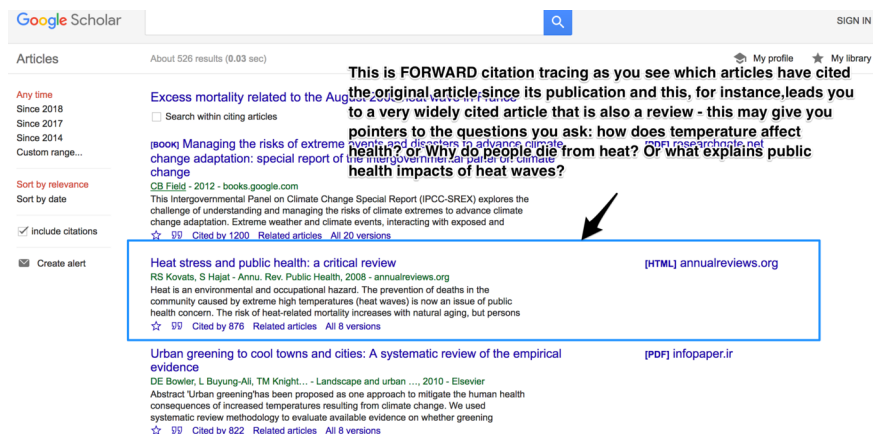
Figure 4: Now you discover even a more highly cited article and one which is a review and was published later than the original popular article. You might as well pick up this article to review and continue with the "Forward" citation tracking to see which articles have cited this article in turn and identify what does the latest trend in the literature tells you. We will leave this exercise for you to complete or you can do this with your own research idea.

These are the papers that were published BEFORE this paper was written, so if you search these papers, you will be running a BACKWARD CITATION TRACING

References

1. Applegate WB, Runyan JW, Jr., Brasfield L, et al. (1981) Analysis of the 1980 heat wave in Memphis. J Am Geriatr Soc.29(8):337-42
2. Bark N (1998) Deaths of psychiatric patients during heat waves. Psychiatr Serv.49(8):1088-90
3. Basu R, Samet JM (2002) Relation between elevated ambient temperature and mortality: a review of the epidemiologic evidence. Epidemiol Rev.24(2):190-202
4. Besancenot J (2002) Vagues de chaleur et mortalité dans les grandes agglomérations urbaines. Environnement, risques et santé.4(1):229-240
5. Braga AL, Zanobetti A, Schwartz J (2002) The effect of weather on respiratory and cardiovascular deaths in 12 U.S. cities. Environ Health Perspect.110(9):859-63
6. Buechley RW, Van Bruggen J, Truppi LE (1972) Heat island equals death island? Environ Res.5(1):85-92
7. Clarke JF (1972) Some effects of the urban structure on heat mortality. Environ Res.5(1):93-104
8. Conti S, Meli P, Minelli G, et al. (2005) Epidemiologic study of mortality during the Summer 2003 heat wave in Italy. Environ Res.98(3):390-9
9. Dessai S (2002) Heat stress and mortality in Lisbon part I. model construction and validation. Int J Biometeorol.47(1):6-12

Figure 5: So after picking up articles that are influenced by your article that got the highest number of citations, you find or identify which articles influence ₋this₋ article and go through the reference list of the article which you refer to as "seminal article". Check if you have read this article before or how the main article (your current article) is citing them or how they have used the information from the articles they have cited in the reference lists and choose and read a few articles yourself

you know that you have got a total pool of resources that you will read and cite from in the first place. So let's say we have run this exercise and we have got to the following article that we will now read:

Li M, Gu S, Bi P, Yang J, Liu Q. Heat waves and morbidity: current knowledge and further direction-a comprehensive literature review. International journal of environmental research and public health. 2015 May 18;12(5):5256-83.

What we will now do is to read this article and abstract information and create a spreadsheet of summarising these articles. But before we do that we need to read the paper closely and construct a standard form to find out hte final conclusion that the article has and its standard form of listing of arguments. This is step 2, so

Step 2. Review the articles and organise the papers that you read

When you do that, you create a spreadsheet with the following headings:

- Citation of the paper
- Synopsis
- Conclusion
- Reasoning
- Evidence
- Your analysis
- Questions to go forward
- Quotations

For the Li et.al. (2015) paper (Li et al., 2015), let's construct this spreadsheet. You can either construct this directly using a spreadsheet such as Excel or you can write this as a free text and then develop your spreadsheet. You can use a citation management system such as Endnote to write the notes on the paper using this format.

How to write a research proposal: what elements to include

Begin with a precise account of the health phenomena that needs explanation. – the background

Discuss the significance of knowing this.

Discuss or Start with an explanation. – Then develop a theory and from the theory build hypotheses

Describe how you will collect data to test your theory.

(if you have analysed some preliminary data then) Describe how your data supports your line of explanation

Acknowledge and list what other explanations might be possible to account for the phenomena you observed

Discuss the limitations of your research approach

Call to Action

References

The Campbell Collaboration. https://campbellcollaboration.org/, a. URL https://campbellcollaboration.org/. Accessed on Sat, February 24, 2018.

Cochrane Collaboration. http://www.cochrane.org/, b. URL http://www.cochrane.org/. Accessed on Sat, February 24, 2018.

OpenEpi Menu. http://www.openepi.com/Menu/OE$_M$enu.htm.URL. Accessed on Fri, February 23, 2018.

The Problem of Induction (Stanford Encyclopedia of Philosophy). https://plato.stanford.edu/entries/induction-problem/. URL https://plato.stanford.edu/entries/induction-problem/. Accessed on Fri, February 23, 2018.

Austin Bradford Hill. The environment and disease: association or causation? In *Proc Royal Soc Med*, volume 58, pages 295–300, 1965.

Tom Chatfield. *Critical Thinking: Your Guide to Effective Argument, Successful Analysis and Independent Study*. Sage, 2017.

Steffen Ducheyne. JS Mill's Canons of Induction: from true causes to provisional ones. *History and Philosophy of Logic*, 29(4):361–376, 2008.

Annunziata Faustini, Andrew J Hall, and Carlo A Perucci. Risk factors for multidrug resistant tuberculosis in Europe: a systematic review. *Thorax*, 61(2):158–163, 2006.

Alvan R Feinstein and Ralph I Horwitz. Problems in the "evidence" of "evidence-based medicine". *The American journal of medicine*, 103(6):529–535, 1997.

Andrew Gelman. Commentary: P values and statistical practice. *Epidemiology*, 24(1):69–72, 2013.

Reina Haque, DN Guha Mazumder, Sambit Samanta, Nilima Ghosh, David Kalman, Meera M Smith, Soma Mitra, Amal Santra, Sarbari Lahiri, Subhankar Das, et al. Arsenic in drinking water and skin lesions: dose-response data from West Bengal, India. *Epidemiology*, 14(2):174–182, 2003.

Mengmeng Li, Shaohua Gu, Peng Bi, Jun Yang, and Qiyong Liu. Heat waves and morbidity: current knowledge and further direction-a comprehensive literature review. *International journal of environmental research and public health*, 12(5):5256–5283, 2015.

Karen A Matthews, Sybil L Crawford, Claudia U Chae, Susan A Everson-Rose, Mary Fran Sowers, Barbara Sternfeld, and Kim Sutton-Tyrrell. Are changes in cardiovascular disease risk factors in midlife women due to chronological aging or to the menopausal transition? *Journal of the American College of Cardiology*, 54 (25):2366–2373, 2009.

Joseph B McLaughlin, Angelo DePaola, Cheryl A Bopp, Karen A Martinek, Nancy P Napolilli, Christine G Allison, Shelley L Murray, Eric C Thompson, Michele M Bird, and John P Middaugh. Outbreak of Vibrio parahaemolyticus gastroenteritis associated with Alaskan oysters. *New England Journal of Medicine*, 353 (14):1463–1470, 2005.

Judith K Ockene. Smoking among women across the life span: Prevalence, interventions, and implications for cessation research. *Annals of Behavioral Medicine*, 15(2-3):135–148, 1993.

Carl V Phillips and Karen J Goodman. Causal criteria and counterfactuals; nothing more (or less) than scientific common sense. *Emerging themes in epidemiology*, 3(1):5, 2006.

Karl Popper. *Conjectures and refutations: The growth of scientific knowledge.* routledge, 2014.

Kenneth J Rothman and Sander Greenland. Causation and causal inference in epidemiology. *American journal of public health*, 95(S1):S144–S150, 2005a.

Kenneth J Rothman and Sander Greenland. Causation and causal inference in epidemiology. *American journal of public health*, 95(S1):S144–S150, 2005b.

Allan H Smith, Mario Goycolea, Reina Haque, and Mary Lou Biggs. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *American journal of epidemiology*, 147(7):660–669, 1998.

Nassim Nicholas Taleb. Black swans and the domains of statistics. *The American Statistician*, 61(3): 198–200, 2007.

Ljubomir Todorović. Original (scientific) paper: The IMRaD layout. *Archive of Oncology*, 11(3):203–205, 2003.

Jan P Vandenbroucke, HM Eelkman Rooda, and H Beukers. Who made John Snow a hero? *American Journal of Epidemiology*, 133(10):967–973, 1991.

Tyler Vigen. *Spurious correlations.* Hachette Books, 2015.

Paolo Vineis and David Kriebel. Causal models in epidemiology: past inheritance and genetic future. *Environmental Health*, 5(1):21, 2006.

Wuxiang Xie, Gang Li, Dong Zhao, Xueqin Xie, Zaihua Wei, Wei Wang, Miao Wang, Guoxing Li, Wanru Liu, Jiayi Sun, et al. Relationship between fine particulate air pollution and ischaemic heart disease morbidity and mortality. *Heart*, pages heartjnl–2014, 2014.