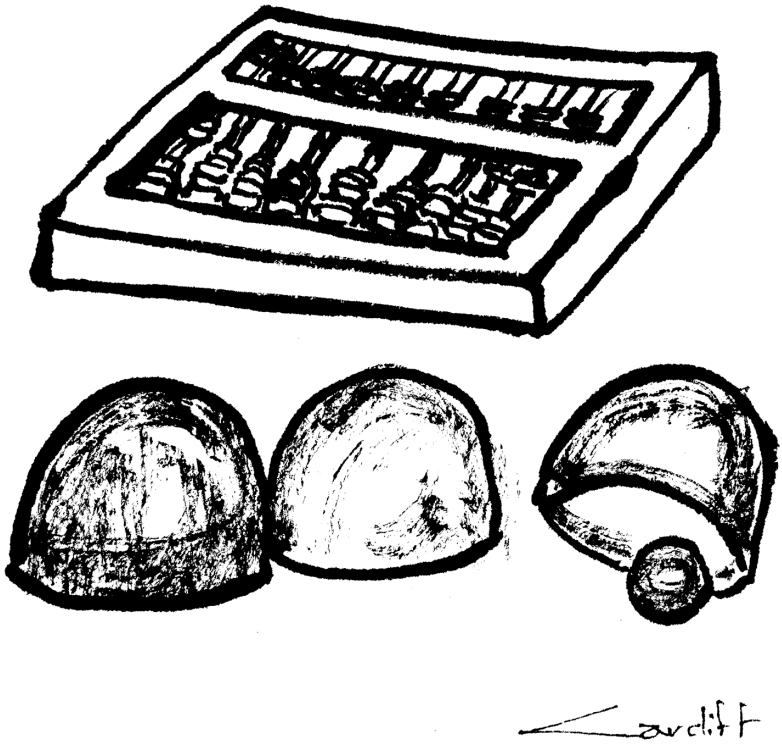
1. The Proposition's Restrictions



To convey meaning

Context demands

The audience knows

And truly understands

Propositional restrictions refer to the characteristics of the problem being studied (i.e., the proposition) in combination with the characteristics of the available environment. These characteristics limit researchers' abilities to derive results that stand the test of time—and many other researchers working on the same or a similar topic. These limits exist because certain things can be done, and others cannot (e.g., see the section on human participants); and certain things are known, and others are not (i.e., researchers never seem to be able to acquire all of the data on their "wish list"). Even things that researchers believe they know are rarely known with certainty or optimal precision. The result is that "perfect research" is most often an oxymoron, as has been seen countless times through the conflicting results touted in the media. These conflicting results, especially when published in a peer-reviewed journal, are not often a sign of poor research. Rather, they are the natural consequence of propositional restrictions, along with other restrictions detailed in later chapters.

The restrictions covered in this chapter are as follows:

a. Questions: the research questions driving the project. Good questions can lead to good research, but poor questions seldom do.

b. Reliability: repeatability or consistency. Whether the issue is test results (e.g., medical, education, etc.) or supposedly the same data extracted from two sources or in two different ways, the results should be about the same. Two "identical" blood tests from different labs should yield the same results from a split blood sample, as should two different IQ tests. Notice, no mention is being made of being correct in any aspect of reliability—simply repeatability or consistency.

c. Validity: the extent to which the available data reflect the characteristics thought to be the ones being studied; the intersection of intent with process. Here is where correctness matters. Repeatability is not enough. Everyone can be wrong, as the history of science has shown.

d. Generalizability: the extent to which research results can be trusted to be accurate for a parent population from which samples were derived. When entire populations are used, generalizability is not a problem.

e. Assumptions: the conditions that are believed to be true without specific evidence. Although inevitable, assumptions are often handled by simply listing them in the research report.

f. Bias: generally, the unconscious prejudicing of the study through researchers' preconceptions or through methodological flaws in the research. Personal bias can be harder to overcome than methodological flaws because it tends to be less apparent to an independent reader.

g. Confounds: the characteristics that might actually be responsible for the results but were not accommodated in the research. Confounds are a major contributor—some would say *the* major contributor—to research results being reversed or greatly modified over time.

The remaining sections in this chapter each discuss these propositional restrictions in turn and how they are accommodated by a high school principal, a director of public health at the state level, and a professor of sociology. Through the examples listed throughout these sections, it becomes clear that research, as it is conducted, is very different from what might be thought. It is most often confined to more limited and poorer quality information than would be optimal for lasting results. These issues set the stage for the fragility of results that is commonly the unspoken hallmark of research.

1a. Questions

Facts don't quite fit
Not sufficiently pat
A revised look
Might answer that

Good research questions are a hallmark of good research. The reason is simple: Research questions motivate and define practically all ensuing aspects of a research design. Good research questions have well-defined terms and are objective, concrete, and answerable within the available resources.

The high school principal wants to know the extent to which extracurricular activities have an impact on grades. His superintendent has asked that the expense for the activities be justified or else the activities might be eliminated from a budget that seems to shrink every year. Although the principal genuinely believes in the value of these activities, his belief alone will not be sufficient to ensure their continuation.

In designing his research, he first drafted his question as, "Do students taking extracurricular activities get higher grades?" Upon reflection, he noticed that different types of students are engaged in extracurricular activities, and even these types of students differ according to the activities. For example, students from poorer families often have to work after school and do not attend extracurricular activities at all. Students on the math team seem quite different from stagehands for a school play in their aptitude for math. Furthermore, few people would assume that participation in a school play would

be as effective at increasing math grades as would a math club. A good research question should be able to accommodate these issues without becoming overly long or complicated.

The principal revised his question to ask, "Is the addition of academic extracurricular activities associated with increased grades in associated courses?" The question has become better by being more restricted, but differences in students who are able to avail themselves of the opportunity to become involved in the relevant extracurricular activities must still be accommodated. Furthermore, the research results must also be adjusted for the initial achievement of the students (i.e., before becoming involved in the extracurricular activities). Nonetheless, the research question has become better by becoming more restricted. More restrictions will be added as other aspects of research are addressed.

The director of public health wants to know if public clinics are providing a substantial portion of the childhood immunizations that do not appear in her data. Unlike private physicians, clinics are not required to publicly report childhood immunizations in the state, and the legislature is reluctant to change the law. The governor wants to know the extent to which a related public health epidemic is plausible, given the numbers of children without a record of being immunized.

In designing her research, she originally asked the question "Do clinics immunize children?" Upon reflection, she sees that the stated question suggests a "yes" or "no" answer as the result. She is really interested in the extent to which children are being immunized in clinics. Her question becomes a bit more complicated: "What percentage of the childhood population receives immunizations from clinics, which do not publicly report the information?"

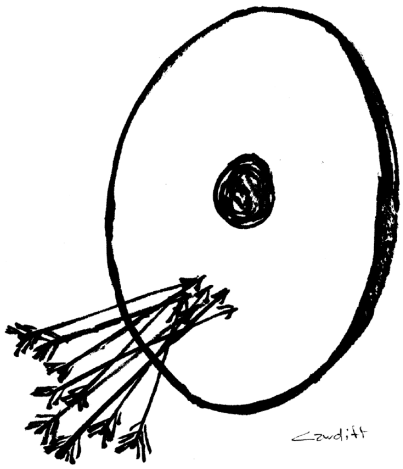
The data for answering her revised question likely will be difficult to obtain. Because the clinics are not required to publicly report the information, it is likely kept (if at all) in a manner that would be burdensome to retrieve. Yet the precision required of the project would likely allow for other methods to suffice. For example, clinics do need to track expenses and inventory. By asking for copies of the immunization shipment receipts and current inventories, estimates of the numbers of children immunized could be calculated by the number of doses no longer in inventory. The estimates would be less

precise where multiple doses per child are needed, but the overall pattern of findings would likely be sufficient for her purpose.

The professor wants to know if matrilineal cultures have more equal rights for women than patrilineal cultures. Her personal interests lie with two west African cultures, and although she has been unable to secure funding for fieldwork abroad, she has arranged instead to meet with groups of immigrants in a nearby major metropolitan area. She has reliable access to several representatives from both west African ethnic groups. Furthermore, she can eat dinner at home with her family most evenings, so she is not entirely displeased with the research restriction of her not being able to travel to Africa at this time. Her original question was, "Do matrilineal cultures have more equal rights than patrilineal cultures?" After discussions with her colleagues and an assessment of her resources, she modified her question to be, "Do west African matrilineal cultures currently have more equal rights than west African patrilineal cultures?" The term *equal rights* is somewhat vague, but it appears to be well understood by her subjects to mean equal rights under the law, as well as by a cultural respect that is demonstrated by courteous behavior toward women.

Matrilineal cultures vary widely in the importance placed on females, and it is possible that our researcher will need to narrow her methodology to case studies of the two specific cultures she is researching. Nonetheless, some of her participants seem knowledgeable about their neighboring cultures, and our researcher is hoping to draw the broadest defensible generalizations from her project. She is reasonably confident that she can answer her questions in a scholarly manner without more costly fieldwork in Africa, at least for now.

1b. Reliability



We all agree
Yet can be wrong
Reliability without validity
Is a very sad song

Reliability is the extent to which different methods or people would arrive at the same data or results. On the surface, it would seem to be a critical aspect of data and results—and it is—when the data and the results are sufficiently correct. Unfortunately, correctness is a characteristic of validity, not reliability. Therefore, although the common conception of reliability is somewhat overrated, it forms the basis for validity, which is critical to good research. The reason is that validity cannot exist without reliability.

The question then becomes, "Why do researchers care about reliability?" The answer is that once something is repeatable, it is much easier to retarget or refocus than if it were not repeatable. For example, darts players first learn to throw darts with an identical motion time after time so the darts land in about the same spot on the board with practice (i.e., reliability). Only then do they shift where the darts will land (i.e., validity, when hit). Simply throwing darts at a board rarely results in a skilled player and most often serves to increase the randomness of the results (i.e., less reliability and validity).

The high school principal has two sources of data for achievement, because the older hard copy system was recently replaced

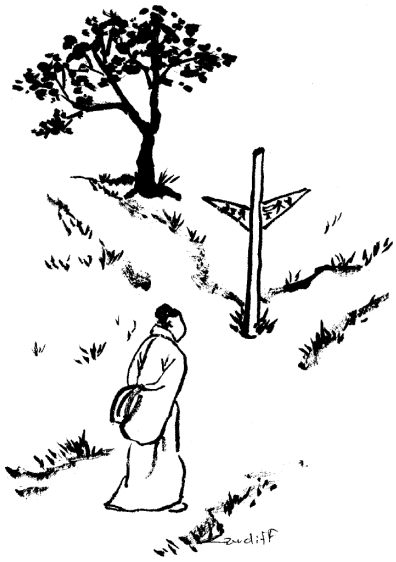
by an electronic system. His first thought is to test the extent to which grades were properly entered into the electronic system (i.e., the reliability of the electronic system). To compare the results of the two systems, the principal met with the district statistician to discuss sample size requirements. The sampling was conducted, grades were abstracted from the hard copy records, and a comparison was prepared. Much to the principal's surprise, only 97.8% of the grades matched. Some were off by more than a letter grade. Although 97.8% might sound like a high proportion (i.e., high reliability due to the high repeatability), the principal noted that it meant that approximately 264 grades were wrong in the population of grades, from a school with approximately 1,000 students. For most social science research, 97.8% reliability would be grounds for a celebration, but not for the principal. He had all of the hard copy grades checked, and those that did not agree were re-entered. A second check on the system revealed the reliability to be 99.7%, which the principal could accept.

The director of public health is willing to accept a far lower standard for reliability, in this case by necessity. The research requested by the governor does not need a precise answer and is really based on judgment coming from information that can be gathered. For example, her intended method for calculating doses of vaccine delivered to children (i.e., clinic receipts minus serum on hand) does not accurately account for some vaccines that require multiple doses. She knows from previous research that approximately 85% of children in her state who receive a single dose from a multiple protocol also receive the follow-up doses, so she can generate some estimates. Yet she also knows that vaccine sometimes expires without being used. Under that condition, state law requires the vaccine to be properly disposed of but does not require a record of the amount. The director of public health will also need to estimate clinic spoilage. These conditions require her acceptance of moderate reliability for the results—at best.

The professor has performed extensive reviews of historical documents on the subject, and she is conducting face-to-face interviews as well as observing the participants. She has arranged to live with two west African families, one matrilineal and one patrilineal, each

for a week during her vacations. Qualitative research is often quite subjective, and even though she is attempting to be as objective as possible, she knows her small study may not be very reliable. She accommodates this by disclosing her restrictions and theoretical assumptions early in her writing. She also plans to focus on her personal interactions with each ethnic group, as she is an attractive young African American woman. The extent to which each group responds to her will be detailed in her findings.

1c. Validity



Certainty conferred
From a human source
Often proves false
When truth runs its course

In a very real sense, validity in research is the result of the intersection of our *intent* with the *process* of its implementation. Researchers believe that they know what they want to measure but often find that their available measures are somewhat compromised by being a blend of what they want to measure and something else. For example, when researchers want to know the impact of a certain drug on the course of a disease, they might be measuring not only the impact of the drug but also the impact of the potentially different lifestyle of people who would volunteer for a clinical trial, even when the volunteers are randomly divided into experimental and control groups. The reason is that the drug might work differently for the volunteers (who would be in a clinical trial and might also be changing their diets, exercising more, meditating, etc.) than for the general population. Random selection into experimental and control groups decreases this volunteer problem but does not eliminate it. The reason is that the drug might interact with one or more of the characteristics that are associated with being a volunteer.

For this example, the assumption is made that the drug is equally as effective as some unknown combination of volunteers' characteristics that (a) differ somewhat from the general population and (b) relieve the clinical condition under study. A second assumption is also made that both these unknown characteristics and the drug can only mitigate the symptoms to about the same degree. Under that assumption, researchers might find very little or no effect from the drug. Yet nonvolunteers (who do not have the critical volunteer characteristics) might have greatly benefited from the drug. The scenario might sound unlikely, but it is the unlikely scenarios that are less often explored, opening the door for compromised validity.

Can complete validity ever exist? Certainly, but complete validity generally requires a very restricted perspective. For example, the statement that a mummified body is dead would seem to be perfectly valid. Yet the statement assumes that the discussion is about the person who used to live in that body. There could still be copious amounts of living organisms within that body, meaning that the body is not completely devoid of life. The life that is still there could easily contaminate other experiments placed too close to the body. So, is it really dead?

Measurement theory combined with vast amounts of supporting literature also shows that perfect validity is a very elusive goal. For example, gender is often regarded as a variable that requires only two response categories to be completely valid: female and male—as evidenced by there being only two types of public restrooms (not including those labeled family or handicapped). Anatomists have long known that physical manifestations of gender form a continuum that is strongly bimodal at its limits (i.e., commonly accepted and mutually exclusive female or male traits), but that it is not insignificant between the typical examples that anchor the ends of the continuum. Intersex individuals (those with a blend of male and female characteristics) have been captured in art and sculpture for more than 1,000 years. If research involving gender ignores this issue, it is easy to see that the measure lacks perfect validity. How might the measure capture the full validity of the gender continuum? Simply adding a third category would suggest that all intersex individuals have substantively identical biological characteristics or features, which is also not true. The physical manifestations of gender form a

continuum that lacks a perceived and well-accepted need for a scale. The result is that many intersex individuals feel disenfranchised when they must choose a restroom or are forced to choose between female and male to describe themselves. If this type of validity issue exists for a trait that is supposedly as easy to distinguish as gender and is very real for actual people, it becomes clearer why validity is an issue when assessing psychological or other less supposedly obvious traits or characteristics that often must be inferred, such as assessing academic knowledge from a relatively small sample of questions.

Physical sciences, such as physics, are not immune to the issue of compromised validity in measurement, as the Uncertainty Principle demonstrates. The very act of measuring something in the physical sciences has long been known to slightly alter the object of the measurement. The measurements might be very close to fully valid, but "close" is not perfect.

Ethnographic research, which often involves a holistic approach to validity through the simultaneous synthesis of many aspects of the issue under study, can be invisibly subject to the impact of pervasive conditioning. Pervasive conditioning is the process by which our lifelong experiences cause us to assume certain things are true when, in fact, they might not be. Pervasive conditioning exists because our experiences are perceptual. The observer, the act of observing, and the object of the observation become intertwined in fundamental ways—on purpose. Although this type of research often reaches conclusions with exemplary explanatory and predictive validity, it might not. For example, ethnographic researchers might want to explore the impact of psychological counseling on depression. Watching the process and seeing that many more patients are improving than not, the conclusion might be drawn that the counseling appeared to be an effective treatment modality. Yet if symptoms of depression tended to resolve themselves with time, counseling could actually be making the condition worse but invisibly so to the researchers. The patients' depression, overall, might possibly have resolved more quickly without treatment, but would the resolution have been as long lasting? Ethnographic research seeks to limit the impact of these types of issues through a long-standing awareness of them and efforts to mitigate them whenever possible.

Even the term *validity* is ambiguous because of the range of plausible research questions. For example, is a measure intended to explain the past, present, or future? Is a relationship between two measures viewed as circumstantial, causal, or both, mitigated by a third variable? Is the researcher seeking to generalize the results to a parent population, while acknowledging the difficulty in acquiring a truly representative sample from most large-scale populations?

On balance, three questions must be asked when assessing most forms of validity. First, to what extent is the measure (or the result) sufficiently valid for its intended purpose? Second, how well can the answer to the first question be evidenced? Third, is the risk of the results being plausibly wrong or potentially harming others worth the information that is provided by the research?

The third question forms the basis for appropriate self-censorship in research because of the inherently questionable aspects of a measure's validity. At some extent, in the continuum of validity, the research answers can become worse than useless; they can become harmful. For example, assume that researchers who were piloting a novel parenting method found that it resulted in children with not only increased adaptiveness but also an increased tendency toward aggression. The researchers might decide to withhold their results until they could disentangle the portion of the parenting method that was associated with increased adaptiveness from the portion associated with increased aggression. At that point, the revised method could be assessed. If the increased tendency toward aggression were greatly mitigated by the newer approach, the researchers might feel comfortable releasing news of the method, along with an appropriate warning on the still existing, but greatly lessened, aggressiveness issue. A more sobering probability is that the researchers seeking a parenting method that increased adaptiveness might not have also thought to test for increased aggressiveness.

Nonetheless, two aspects of validity become clear once it is understood that validity ultimately relies on logic or judgment, at least to some extent. First, there is an arbitrary nature to validity. One only has to examine the performances that were considered perfection in the diving competitions in the Olympic Games over the decades to see the evolving nature of a perfect score—the fully valid exhibition

of the skill. Performances that took gold medals many years ago would often fail to qualify for a national team today. Why? Because the competitions' judges long ago did not believe that people were capable of safely performing some of today's more complicated maneuvers. Their pervasive conditioning led them to believe that the best that they saw was the best that existed—or perhaps could exist. The distinction is not always clear, but it speaks to the heart of validity. When is something exactly what it should be? How should it be compared with something else that is exactly as it should be but is much harder to do? In today's Olympic Games, the degree of difficulty for a dive is accommodated through its maximum allowable scoring, but a remnant of this issue still exists.

This remnant leads to the second aspect of validity that derives from it ultimately resting on logic or judgment—an appearance of arrogance that is associated with the authority to decide the extent to which a measurement (or Olympic performance) is valid. What makes someone qualified to pass judgment on validity or even to decide how it will be measured? How can any individual escape lifelong pervasive conditioning to be able to employ fully objective judgment? Is validity hopelessly confounded with temporal events? Is the nature of validity inexorably linked to one's environment and place in time? Should or can enduring validity be a special type of validity that is not necessary for some types of research? If so, for how long, and how do we know? Issues of reliability suddenly seem so simple by comparison to issues of validity, which can be quite humbling. In time, some researchers become somewhat cavalier to the difference. When they do, they often find that relaxed attitude to be a big mistake.

The principal knows that differences in student grades often involve a constellation of characteristics interacting with their environment. He knows that his question is not only simultaneously simplistic on its surface, but also far too complicated to support more than a tentative answer, regardless of the logic suggesting that academic clubs are designed to promote achievement in their subject matter. He knows that, although the association between club content and course content is logically sound and pedagogically defensible, his superintendent will want to see quantitative results in addition to qualitative reasoning.

In considering his approach, the principal further reflected on the nature of validity because it would have an impact on both his qualitative reasoning and research into the clubs' and courses' contents and interactions with students, along with his quantitative results. From a qualitative perspective, he understood that the clubs were developed with charters that specifically addressed course content. He saw students participating in club events that evidenced course content. Teachers stated that the students in the club increased their knowledge of specific course content more than students who did not participate in content-related clubs, if for no other reason than increased practice. The principal was fairly certain that his qualitative case was reasonably strong, despite the plausible caveats on the lack of independence among the observers and the events.

When the principal thought about the ways in which plausible issues with validity could affect his results, he reworded his research question a second time. His question was, "Is the addition of academic extracurricular activities associated with increased grades in associated courses?" It became, "To what extent does the addition of academic extracurricular activities appear to be associated with increased grades in associated courses?" Now the question more clearly implies not only the need for a quantitative answer, but it also suggests that his answer will fall along a continuum, rather than the yes/no answer previously implied. Furthermore, his new question also includes a sense that he might not be able to arrive at a definitive answer—a personal disposition that is more aligned with a research perspective and speaks to the somewhat elusive nature of validity.

The director of public health knows that she will have data with at least somewhat compromised validity. Issues involving multiple doses for many types of vaccines and the disposal of unused serum impose the need for the acceptance of more ambiguity in her results than for much of the other research that she has previously done. This ambiguity translates to decreased validity. She will need to decide if the amount of truth in her answer sufficiently outweighs the uncertainty in it to be justified. The validity of her results would be somewhat compromised by the uncertainty in her measures, but she feels that there is enough "truth" that she can capture to yield a reasonable answer. Therefore, she will appropriately qualify her report's

results to reflect her perception of the degree of validity of her data and refer to that report if her results are ever taken out of context—a somewhat persistent concern for many people who write documents within the public domain, such as hers.

The professor's research is much more qualitative in nature compared to our other two researchers' projects. Although there is an abundance of data on the status of women in each culture, her question is intentionally vague enough to include input from her fieldwork. Nonetheless, the anecdotal information that she gathers will need to agree substantially with the facts garnered from the international data on education, employment, health, and laws. One key fact she had already found is that women own 50% of the land in the matrilineal culture she is studying, compared to 2% for the neighboring patrilineal culture, which is the second culture she will survey.

Her qualitative research is studying a slice in time and is concerned with topics that do not fit into preexisting categories. Validity is couched in terms of the many restrictions on the complexities inherent in human interactions. Her study participants believe that they are offering valid opinions and true reflections of their own experiences. Yet their own situations could distort their perceptions sufficiently to make a general statement of validity impossible to justify. The professor knows this and seeks to gather information from sources with varied backgrounds within each culture. She will be interviewing new immigrants with refugee status, as well as college professors and successful business owners. Her past research has taught her that she will be hearing very different stories from each group, and accommodating those will be a challenge.

1d. Generalizability

The importance of context
Cannot be overstated
For results to generalize
Situations must be related

The results of research are said to be generalizable when they apply to a larger group than was studied. For research results to be generalizable, one of two conditions must exist. First, the sample used for the study must be a sufficiently valid subset of the overall population, or, second and alternatively, the group to which the results are being generalized must be shown to be substantively equivalent to the studied population for all traits that could impact the topic of the study. Although deriving an appropriately representative sample is conceptually easier and most often the method employed, it is not devoid of threats to the validity of the generalizability.

An example often used is that of researchers who want to survey a sample from a town's population. How would it be done? Phone lists would not capture certain types of individuals who tend to have unlisted phone numbers. Property lists would not capture renters. Even a complete list of all residents in the town would miss homeless people. Such a list could also include units in hotels and motels—some of the residents of which are permanent, although many others are not.

At some level, the idea of completely generalizable research results often disintegrates. The task of researchers is to be both careful and thorough in their descriptions of the extent to which generalizing might be problematic for their results. This issue is at the heart of the sections of peer-reviewed journals that speak to the

population, the sample description and sampling methodology, and some of the limitations that are discussed. Although many readers skim over the Limitations sections thinking that the issues are little more than technical verbiage, these sections spell out the extent to which the knowledge generated has been restricted by the various conditions under which that knowledge was generated.

The principal views his populations as relatively small, and his electronic records system has been tested for reliability with regard to student grades. Upon some reflection, the principal realized that agreement (i.e., repeatability) between the two sources of grades met the definition of reliability, but one of those sources was the permanent student record. The contents of that record are considered historical truth, even when records are recent. The principal realized that he had actually verified the validity of his electronic records systems when he had originally thought that only the reliability had been assessed. Then the principal remembered that only the validity of the students' grades had been validated and not any of the other information in the electronic record.

The principal now faces the choice of validating the other types of information that he will be using to try to adjust for the types of differences in students that could be related to both attendance at academic clubs and improved achievement in the related courses. If he validates the other types of information that he will be using, he will not need to use samples but could use all of his student data. Given that the impact of clubs might be relatively small, he wants all of the statistical *power* to find a difference, if one is there—which he knows comes with larger numbers of students in his study. An additional advantage is that generalizability is not an issue when the entire population is used, unless the principal wants to generalize over time to a future group of students.

The appropriateness of using a population as a sample in time to compare or predict results often draws a mixed response from both research methodologists and statisticians. Except for truly longitudinal studies, little, if anything, is generally done to show the comparability of the two populations over time that are being compared. Changes in the measures could be due, at least in part, to changes in the population, which is rarely the topic of a study. In short, statistics are

designed to explore differences that involve samples, not populations. For that reason, many research methodologists are somewhat reluctant to design studies that use statistical reasoning with populations. Nonetheless, the use of populations across time as if they were samples in time is a staple of public health research. For public health, changes in the population do not change their goal of mitigating disease, especially through process changes that have an impact on entire populations.

The director of public health has the electronic records for every affected individual in her state. Although she could conduct much of her research with entire populations, she wants to examine the uncertainties surrounding multiple vaccine doses and disposal. To do so, she will need to examine the records from a sample of clinics. She understands that the sampling is not straightforward. She will have samples of both vaccine shipments/disposals and beneficiaries, both of which are contained within the processes and procedures of clinics (a situation called *nesting* in research). Further complicating her acquiring anything close to a random sample, some of the clinics are fully independent whereas others are part of a chain with shared practices. Trying quantitatively to understand dosing and disposal will require the assistance of the department's statistician for both the sampling and the proper aggregation of the resulting information (data become information when placed in a context).

The professor knows that her results will need all of their restrictions carefully noted in order for them to be generalizable. Nonetheless, her research is one small piece of the larger puzzle that is human nature. Some of the pieces already in this puzzle include well-accepted findings, such as the following: Women who attend school can be generalized to have more opportunities for employment than women who do not receive a formal education, and better employment can be generalized to increased access to health care and decreased exposure to dangerous living conditions. To the professor, understanding the role that tracing a lineage through the mother (instead of the father) plays in empowering women seems an interesting piece to the puzzle.

1e. Assumptions

A critical piece
To fill in the gaps
That unknown fault
Where mountains collapse

Assumptions can be (and often are) defined as the specific data and situational requirements for statistical and methodological techniques, respectively, to be valid. Statistical assumptions are listed in many statistics texts and are often liberally sprinkled with Greek letters. Although this category of assumptions is important, its handling is generally best left to statisticians and research methodologists when technical issues need consideration. The outcome of minor violations of statistical and methodological assumptions for many, if not most, studies is marginally biased results that do not invalidate the substantive findings of the research. The reason is that substantive knowledge from researchers in conjunction with review, oversight, and publishing requirements generally results in reasonably sound statistics and research methodology.

The bigger threat from assumptions for the validity of research stems from our unknown assumptions. These assumptions arise from a lifetime of experience with the way we interact with the world. For example, if a dozen or so American men were asked the number of outs in an inning of baseball, most odds-makers would suggest that at least a few of those men would be willing to wager their last

dollar that they knew the correct answer. Yet in many parts of the world, the men would look at each other and ask if any of them knew what an "out" or an "inning" was. Knowledge, actions, reactions, and a host of other traits that might be expected of others could be the result of how we grew up and were conditioned—hence, the term *pervasive conditioning.*

Pervasive conditioning can mask true causes by affecting both the target of the study and whatever researchers believe is responsible for variation in it—and researchers could be definitionally unaware of it. Peer-review processes, international committees, and other structures have been established to try to limit the impact of this type of threat to the validity of research results. Yet the threat from pervasive conditioning for the validity of results can also serve to remind researchers to be somewhat humble with regard to those results. Again, being cavalier about the impact of pervasive conditioning on validity has caused many reputations to be damaged.

The high school principal knows that the conditions of students' self-selecting academic clubs and coming from backgrounds that do not require after-school employment violate the notion that he might be able to look at a random sample of his students. His findings need to be qualified only to apply to the types of students who were in the academic clubs. This part of his assumptions does not trouble him.

Yet the nesting of students within clubs is a difficult situation, both statistically and methodologically. If randomness cannot be assumed, many of the statistical and methodological techniques taught to undergraduates are not valid approaches without some type of special handling. Knowing when he does not know enough, the principal will discuss the various stages of his project with his district's statistician—also the district's research methodologist.

The director of public health will have few, if any, problems with the assumptions underlying the use of her electronic data. Those data are population based, have been used for many years, and are well understood. Nonetheless, her study of multiple dosing and disposal presents problems. The state's clinics are not independently owned and operated, so their processes and procedures are likely shared and cannot be considered as randomly selected examples. The assumption of independence in observations is central to many methods of sampling. Sampling

under the condition of poorly characterized partial dependence will certainly motivate her to involve the district's statistician.

The professor has made assumptions based on her previous fieldwork to western Africa and her more recent review of the current situation. Her study participants have been specifically chosen to challenge these assumptions. She wants to hear all sides of the story. Her contacts with more affluent members of the ethnic groups have led her to believe that women who own land have more security and independence than those who are landless. By including new immigrants in the study, many of whom came from dire circumstances, the professor hopes to challenge her own thoughts about land ownership.

The professor is working with new data, which were gathered by a highly credible international organization. The new report is much more complete than were previous reports, and it serves as the backbone for the quantitative portion of her study. She has discussed her project with others in her department, and she consulted with the college's research center, where a research methodologist reviewed her proposal and found that it was sound. He advised her to continue to scan the current literature for similar articles, as it would not be due diligence to duplicate another person's research. Our professor hopes to publish her work and present her results at a national conference, and she recognizes the value of his advice.

1f. Bias

Our parents before us
Friends and culture, too
Color our vision
Twist our point of view

Just as for assumptions, two types of bias are often considered in research: personal bias and technical bias. Personal biases, generally from our pervasive conditioning, can substantively affect several aspects of complicated studies because the number of judgments and decisions can be enormous. These decisions, all of which are influenced by bias, can greatly influence a study's results. Most researchers and educated consumers of research understand the issue and use structural methods, such as random sampling, whenever possible to avoid many of these judgments and decisions. Limitations sections are often included with results to discuss plausible sources of bias, among other topics.

Technical bias exists when the chosen techniques systematically overstate or understate the "true" results. Many sources of technical bias have been well characterized, such as the misleading nature of using the arithmetic mean as the average of a population's income. In this example, extremely large incomes by relatively few individuals can make the average income appear functionally larger than it really is. An extreme case could be when one individual earns $10 billion, and 999,999 people earn nothing. The arithmetic mean suggests that a typical person in that group earns $10,000. Yet only one individual earns even one dime. Both the median (the point where half are lower and half are higher) and the mode (the most common occurrence) return $0 earnings for the typical person in that group, which is a functionally unbiased estimate of typical earnings compared with the arithmetic mean.

Fortunately, technical bias as a restriction on the basic propositions driving research has been well characterized for most situations, and appropriate techniques are generally employed to mitigate the effects. Nonetheless, examples reach the public with almost alarming frequency. Continuing the previous example, average salaries for various occupations are often presented in the media. Which average is being presented (i.e., mean, median, mode)? How variable are the presented amounts by experience and by geographical location? Who gathered and supplied the data, and might they have an alternative agenda? It quickly becomes clear that the absence of a detailed context results in little or no useful information. The reason is that data require a context to become relevant, and relevance is required for information to be useful.

The principal knows that the population of students in academic clubs is a biased sample of his entire student population. Nonetheless, he also knows that he can accommodate this source of bias by being careful with the strength and generalizability of his results. At best, he expects his results to tentatively support the hypothesis that increased academic gains are associated from club activities. But now he has an additional problem.

The principal just realized that he has very real expectations for his results before embarking on the study. Consciously or not, he knows that his objectivity might be compromised. For this reason, he asks his district methodologist for ongoing oversight as a method of ensuring an unbiased approach to the various stages of the project.

The director of public health does not foresee any problems with technical bias with regard to the portion of her work that used her electronic data. The clinic sampling and resulting data, however, require special statistical handling or would likely show biased results. Moreover, even with special handling, the results are likely to be somewhat biased in unknown ways because of the relative lack of current information on the extent to which ownership and policies overlap across clinics.

The professor acknowledges her personal and professional biases. After 10 years researching women's issues in west Africa, and with the addition of better statistics collected by international organizations, she feels that she knows her outcomes in advance. Nonetheless,

she is trying to maintain a fresh perspective for this study and has included more men and a greater diversity of women than previous studies that she has read.

The situation where a researcher knows there might be biasing technical considerations that cannot be adequately rectified is all too common. This issue has resulted in the Limitations sections normally found in scholarly articles. Through the Limitations section, researchers delineate known and potential sources of bias for their results.

Personal bias is rarely, if ever, discussed in research results because personal bias has no legitimate place in research. Nonetheless, people design, conduct, and analyze the data research studies. The existence of pervasive conditioning alone argues for a more transparent approach to acknowledging and accommodating the issue, as was done by the principal.

1g. Confounds

Evidence confirms
It must be so
Heard later in the halls
We just didn't know

Confounds are the missing ingredients waiting to catch up with every researcher. A confound is a variable that is related to both the dependent variable (the variable of interest) and the independent variable of interest (an explanatory variable) but whose influence has not been ruled out or otherwise accommodated.

Three ways exist to accommodate potential confounds. The first and best method is through the research design, where a strong design mitigates against many common sources of confounds. The second method is through accommodating statistics, where the impact of these variables is mathematically controlled. The third and weakest approach is by appealing to logic and argument: the "it only makes sense that . . ." tactic.

Few journals accept the third technique. Most reputable information outlets (journals, books, national media, etc.) prefer the first method but also will accept strong examples of the second. Publishers know that confounds eventually surface, are resolved, and often overturn research results; so, they are particularly sensitive to research designs that are more prone to plausible confounds, such as those using a nonrandom sampling methodology.

In the end, though, researchers can neither think of nor control for everything that might be responsible for some of the findings that were ascribed to something else. To advance knowledge, researchers must be willing to risk being wrong, while trying hard to avoid it.

Importantly, it is often from showing where others were wrong that knowledge achieves its largest gains. The reason is that others are not restricted by their pervasive conditioning in the exact way that the original researcher was restricted. Because of the uniqueness of what each of us knows, combined with the differences in each of our pervasive conditionings, others can see our research mistakes more easily than we can see them. Knowing this cycle of iterative improvements on others' work, researchers are wise to remain somewhat humble in the presentation of their results.

The principal realizes that the potential list of background and personal characteristics that could be associated with academic club participation could be enormous. The thought that he could control for all of them would be preposterous. Even if he knew exactly which variables would need statistical control and had the required data, he would not have a large enough sample for the basic calculations. The reason is that the needed sample sizes increase rapidly when control variables are added to the equations. Understanding this issue alone is a sufficient reason to maintain a strictly tentative approach to the reported findings.

The director of public health knows that the issues of multiple doses and the lack of complete disposal documentation will make her results somewhat tentative. Yet her work will yield a better estimate of immunization coverage than has been previously available. Although she doubts that she will miss anything major, the issue of confounds often cannot be completely resolved in most research.

The professor is pleased with the new data, which she feels support her hypothesis and add objectivity to her long-held belief that land ownership patterns associated with matrilineal cultures lead to more security and independence for women and their children. Nonetheless, with the complexity inherent in human nature, she anticipates that she will glean some new insights from her fieldwork that might present one or more confounds that she will need to accommodate in some manner.