Foreign Language Short Course

Special Operations Clinical Research Fundamentals

Kate Rocklein Kemplin, DNP, RN, CCEMTP; F. Young Bowling, 18D/18Z, ATP, NRP, BHSc

ABSTRACT

When we do not know a language, we are at the mercy of an interpreter. The same is true for research: Special Operations Forces (SOF) clinicians not actively involved in research initiatives may rely on scientific interpreters, so it is important to speak some of the language personally. For any clinician, using evidence in practice requires a working knowledge of how that evidence was generated from research, which requires an understanding of research science language. Here we review common basics of research science to reinforce concepts and elements of experimental and nonexperimental research.

KEYWORDS: *research; statistics; statistical analysis*

Introduction

Research is a scientific process that validates and refines existing knowledge and generates new knowledge that directly and indirectly influences the delivery of evidence-based practice. The four main areas we address with research are: *Description:* What is it? *Explanation:* What does it explain? *Prediction:* What occurs with it? And *control:* What changes happen when it . . . ? For descriptive research, we might seek to determine the incidence of cellulitis in candidates going through the Q course. Explanatory research would seek to determine if incidence rates of cellulitis differ between months that candidates go through the Q course. Predictive research would seek to determine the incidence of MRSA infection in cellulitis in candidates going through the Q course at different months. Controlled, or experimental, research would seek to determine if chlorhexidine bathing decreased the incidence of MRSA infections in Q-course candidates who had cellulitis during specific months of training.

All research has philosophical underpinnings (which is why degrees in research science are referred to as PhD: doctor of philosophy). Quantitative research is formal, objective, and systematic: structured processes are used to obtain numerical data to answer a research question. For quantitative research, the majority of the underlying philosophy stems from logical positivism, founded in rules of logic and laws, whereas in postpositivism, probability is a form of truth versus cause-and-effect frameworks. As we break the whole into parts, we use reductionism to determine what elements should be examined. This requires objectivity and logical, deductive reasoning.

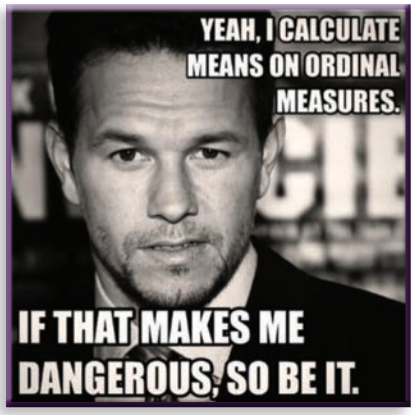
Most quantitative research methods involve descriptive, correlational, quasi-experimental, or experimental research. In descriptive research, we describe the individual(s), phenomena, or group(s) of interest; in correlational research, we are looking for relationships and associations; in quasi-experimental research, we identify causal relationships and their significance to each other, or a combination or associations and causes; and in experimental research, we seek to find the causal relationships between control and experimental or intervention groups. In this article, we will speak mostly to experimental research and touch on nonexperimental research.

Variables

All research involves measuring some kind of variables, or phenomena that we notice in practice. We can measure variables that are naturally occurring (intelligence, height, race, etc.), we can control for certain variables, or we can manipulate them. We have to understand what variables are before we can measure, use, or manipulate them. We need to determine how variables are best used, such in experimental or nonexperimental research. Categorical variables are discrete or qualitative variables and can be nominal, ordinal, or dichotomous; continuous variables are quantitative variables known as interval or ratio variables.

Data are composed of variable measurements, and measurement precision is critical. Interval measurements are the most common type of measurement, such as temperature, and they are measured along a continuum or scale. Ratio measurements possess a relationship to scale, for example, 100kg is twice as much as 50kg, although they do not have negative values—you cannot have negative weight or height (e.g., you cannot be –10kg or –10cm). Ordinal measurements are exactly that: measurements that are in a particular order, that express a magnitude of effects. The Richter scale is an ordinal scale: destructive earthquakes have magnitudes between 5.5 and 8.9, and with every 1-point difference on the scale, magnitude has increased by about 30-fold.

The pain scale from 1 to 10 is an ordinal scale, though it is subjective: what is a 10/10 pain level to one person is a 3/10 pain level to another (see also: differences between women in childbirth and men who have a cold). Nominal variables are variables that have two or more categories but do not have an intrinsic order. For example, we could classify American SOF Medics into SOCMs, 18Ds, and IDC Medics, which would be a nominal variable with three categories. We could call those categories groups or levels as well. Dichotomous variables are nominal variables that have two choices: yes/no, military/civilian, or, historically, male/female.



We assign ordinal or nominal measurements to variables (such as pain level) to operationalize a concept. We likely are not going to give narcotic analgesia to someone who states, "I kind of have a little bit of pain"—we are going to operationalize that vague concept into a rating on a pain scale so we can score it more objectively and communicate those findings to others. We do this to measure emotions, aggression, depression, and various other concepts, such as reasoning and even opinions. Posttraumatic stress disorder (PTSD) checklists and similar items are scales of measurement that provide a numeric score, which we can then analyze statistically through quantitative analyses. Ordinal and nominal measurements are arbitrary, though, and should be interpreted with caution. We use them to operationalize nebulous concepts (depression, suicidality, etc.) into numbers that do not actually represent concrete points. A Likert scale (e.g., 1 = strongly disagree, 5 = strongly agree) gives you a range of responses or levels of agreement: it is subjective and open to interpretation. Scoring a 1 on a Likert scale does not mean you are 5 times less in agreement: it is a number representing a range of reported agreement with a concept.

Operationalization

Despite those caveats about subjectivity, operationalization is supremely important. For example: survivability. Survivability means different things in engineering, medicine, and in the military. Within the military, survivability is composed of several different elements, and survivability even has separate meanings between ground forces and the Navy. In our medical contexts, what does survivability mean? Does it mean you survived your initial trauma and made it to a higher echelon of care? Does it mean that you died within 1 year of injury? Does it mean that you were successfully resuscitated and died within 30 days? To measure survivability, we would have to carefully and systematically operationalize that variable and its determinant components.

Strictly defined, "operationalization" is the process of defining the measurement of phenomena into terms that are understandable empirically. Essentially, we take a fuzzy concept—depression, satisfaction, racism—and distill it into criteria and elements that prove or define its existence. In health care, we may use body mass index (BMI) to measure overall fitness, though anyone who has been tape-tested before a PT test would probably tell you BMI is a poor operationalization of overall physical health or fitness estimations.

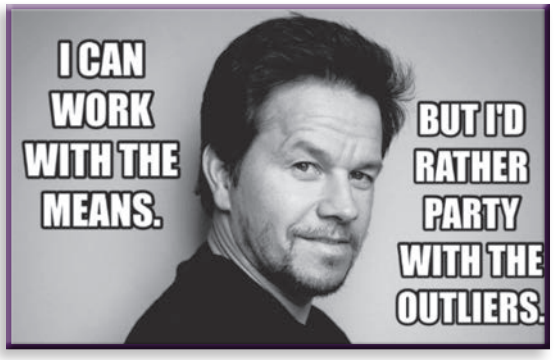
Operationalization deals with complex concepts in attempts to measure their presence. A major flaw is that operationalization assumes that there is a baseline or a uniform measurement. If you were evaluating this article's primary author in terms of loudness, emotion, or use of expletives (the opposite is likely true for the more-reserved second author) with an operationalized scale, the element of "stereotypically feisty Irish Catholic" should probably factor into the baseline and subsequent measurements.

Point being is that operationalization can be subjective and independent to the participant being measured; it requires a lot more than consensus science or expert agreement. If we are going to actually measure concepts like survivability, particularly within SOF, we will have to deconstruct the words and phrases we use to measure that outcome (dependent variable) into component parts that concrete and quantifiable. Some of us might view survivability as being physically alive within 30 days of combat polytrauma, some might see survivability as being functional a year after sustaining trauma, and some might see neither instance as encapsulating or representing survivability. As we continue with, say, prolonged field care (PFC) research and practice development and venture into new areas of practice, operationalizing variables will likely take on a new importance and meaning in different contexts as Medics' scopes of practice develop and advance.

Independent variables (IVs) are those that are manipulated or considered experimental or predict something. Dependent variables (DVs) are the outcomes you are looking for or trying to measure; they "depend" on the independent variable. A basic example would be hours spent studying (IV 1) and intelligence quotient ("IQ"; IV 2): both contribute to, or predict, a score on an examination (DV). In an experiment, we would manipulate an IV, such as hours spent studying, and then expect a change in the DV: the score on the examination. Specifically, we would expect the examination score to be lower if the hours spent studying were fewer, so that is a positive relationship—more hours studying would likely make the examination score rise. In short, the dependent variable literally depends on the independent variable. The expected outcome is what we hypothesize, which essentially is an educated guess. A null hypothesis would read this way: there is no relationship between examination scores and hours spent studying. An alternative hypothesis reads like this: there is a relationship between examination scores and hours spent studying. Let is not forget intelligence, which may affect examination scores as much as hours spent studying. Technically, we cannot change someone's IQ, so that IV is not going to be manipulated. However, we could instruct one group to study *x* amount of hours, which would be a manipulation of an independent variable.

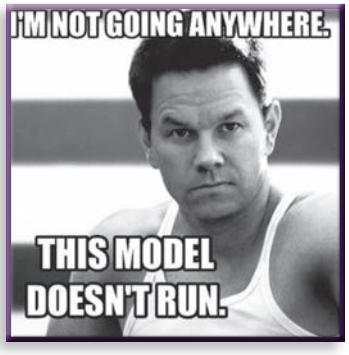
Ways to Be Average

"Measures of central tendency" refer to the average, or center, of a set of numbers. We tend to think of the arithmetic mean as the average, and the mean is admittedly a commonly used calculation in statistical formulas. The big problem with using the mean is the outliers: it is very sensitive to extreme (high or low) scores, especially if your sample is smaller and does not have hundreds of participants. The mode is the number occurring most in a set of numbers. It is a central cluster of the values that appear most. The last measure, and possibly the strangest, is the median. The median is the middle number in a set of values. A lot of salaries and mortgages are reported this way in national data sets, which really does not make sense—we could draw a line in between value 49 and value 50 of a set of 100 numbers, but that does not reflect the true average of the values.



"Data" Is Plural, and "Statistics" Is Not a Dirty Word

To quote John Tukey, hero of boxplots and loved by statistical supernerds: "Statistics is a science using mathematical models as essential tools." We use statistics to explain and interpret data. Without statistics, everything is just conjecture: reporting percentages and raw numbers is grammar-school math, not statistical inference. Data are composed of information expressed as numbers. Data, for our purposes here, refer to quantitative information. As with most endeavors, statistics are only as good as the data put into it: garbage in, garbage out.

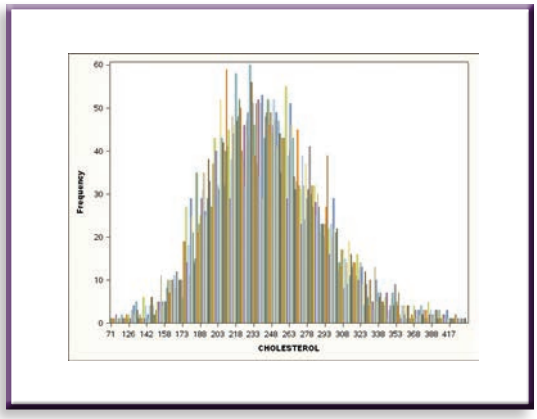


Speaking of garbage, beware of ignoring data reporting and tables in journal articles. Often, readers gloss over figures and tables because when data are presented

formally, we assume that the results are legitimate if they appear legitimate. Those are dangerous assumptions to make. First, beware of "statistical" reports that only give you percentages and raw numbers. It is nice to know how many crics or cutdowns were performed downrange out of *x* amount of patient encounters, but without comparative or correlational analyses, ratio data (percentages) are mathematics, not statistical inferences. Percentages are ratio and frequency data, and little more than that. Second, we have to grasp the difference between clinical significance and statistical significance. If we hack at data long enough with statistical tests, we can make anything statistically significant. Especially hilarious is when investigators emphasize "significant" p-levels when they're reporting statistical tests in which p-levels are essentially irrelevant. [More on that in future articles.]

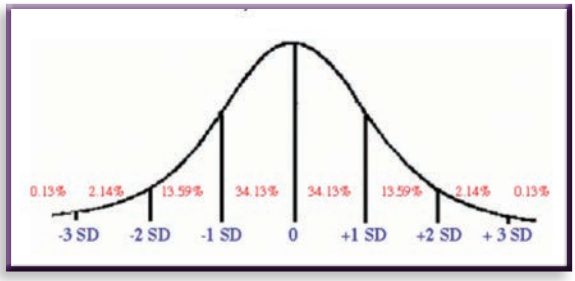
"Normal" Does Not Mean Good, and "Significance" Is Often Irrelevant

Whenever we use inferential statistics, we make some major assumptions: one being that the DV (such as cholesterol levels in a specific group after taking a statin) is going to be normally distributed. "Normal" in this context does not equal "good"—it simply refers to how data are distributed. We see here that the distribution of cholesterol levels in this completely made-up sample is centered around an average score of 220mg/dL. Looking at the curve(s), we see that the cholesterol levels meet assumptions of normal distribution: it is symmetrical in that the left and right halves on each side of the peak (the average) are equal. Just as many scores are higher than the average than are less. It is bell-shaped with the highest point around the average, and it is asymptotic: the values never actually meet the *x*-axis though that is hard to determine just by eyeballing it. In a normal distribution, whether it be cholesterol, age, or test scores, 99.7% of results are within 3 standard deviations (SDs) of the average, or "the area under the curve"; 95.4% of results are within 2 SDs, and 68% are within 1 SD.



Standard Deviations

The SD is the square root of the variance, and the variance is the average of the squared distance that each score is from the mean. So if 220 is the average cholesterol in this group of patients, we would take each individual score and subtract it from the mean, and then square the result. We would calculate the mean of those squared differences and take the square root of that . . . but we would use statistical software to do that for us.

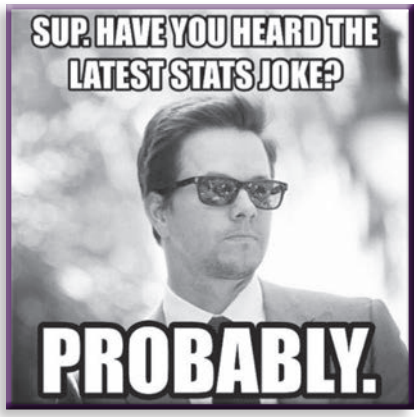


Probability

If the distribution is normal, we can examine the probability that statins lowered cholesterol levels. This is where we see results like $p < .05$, which literally indicates that there is a 5% chance that our results are a fluke. A *p*-value is an estimate of getting results by chance; if it is a small chance, like 5%, then we are assuming the risk that 5 times out of a 100, the results we get (lowered cholesterol levels) have nothing to do with the statin. It does not mean anything more than that. If we compared cholesterol between the control group and statin group, and $p = .03$, it is less than .05 and we would accept that there is a statistically significant difference between the groups. If the cholesterol levels in the statin group were lower, we could infer that the statin lowered the cholesterol levels. Note that that does not mean there is clinical significance. Even if the differences in cholesterol were statistically significant, that does not mean that the clinical outcomes changed or that patients fared better.

Confidence Intervals

If we sample a bunch of people to gather data, we will get a range of values. A confidence interval (CI) is a range of values in which we believe the true value of a parameter (numerical characteristic of a population) is held. The CI is the percentage of intervals that contain



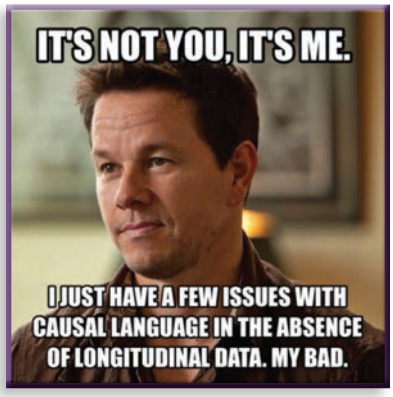
that population parameter we believe we would obtain if we took a lot of measurements from that sample population over and over again. For example, say we randomly choose to measure the average height of every male in SOF. That sample's average height is a point estimate of the population, which is uncertain because we do not know how far the sample average is from the average height in all of SOF. We cannot say with real certainty that the average we calculated represents the entire population. So we use statistical procedures to estimate the population mean from a sample that we believes contains the true value of the population's mean either 95% or 99% of the time. That does not mean we are right 95% to 99% of the time; it means we have used a mathematical model to estimate the true value of a population. If we were to repeatedly make new estimates with identical measurement procedures, the CIs would contain the average of all the estimates 90% of the time. So our estimate results in 90% of the CIs formed containing the true value.

Experimental and Nonexperimental Research

Experimental research is the "true science" we think of when we hear the word "research"—it is considered the most rigorous of all research designs, but it can actually be a lot easier to perform than other types of research, because it is very cut-and-dried. Either you did it right, or you did not: there is very little ambiguity. In experimental research, IVs are manipulated and compared with a control. In true experimental research, we not only manipulate the independent variable; we randomly assign participants to the control and experimental groups. For example, if we wanted to lower serum cholesterol (DV; our outcome variable), we would give one group of patients a statin and another group of equal numbers with similar patients a placebo; to make it "double-blind" we would also hide which medication was the statin and the placebo from the researcher assistants administering the drugs to the groups. Provided that the groups are homogeneous (e.g., similar in ages, ethnic backgrounds, comorbidities, etc.), we can responsibly compare the numerical results (also known as "quantitative" results, because we are counting things) of serum cholesterol after taking a statin for 90 days, versus the group who took a placebo medication. In this example, we could estimate a cause-and-effect relationship that we hypothesized.

Our null hypothesis is that statins have no effect on cholesterol levels. Our alternative hypothesis is that statins lower cholesterol levels. Because we stated an actual direction of the effect of statins (lower cholesterol), the alternative hypothesis is also considered a directional hypothesis—something is increasing or decreasing in relation to the variable we manipulated—in this case, giving people in one group a statin and people in the other group a placebo. When we compare the results of the placebo group with the results of the statin group, we are not going to look at raw numbers alone. Raw numbers do not tell us a whole lot, and they are really ambiguous. Cholesterol is a good example: a total cholesterol level less than 200mg/dL is what we aim for in the United States. So if your cholesterol level is 186 and another person's is 178 after taking a statin, what does that mean, really? Is that a significant difference? Does that 8-point difference actually mean another person's risk of cardiovascular disease is lower than yours? This is where statistical analyses become important.

In nonexperimental research, we do not manipulate the IVs, usually because it is either illegal, impractical, or unethical. We are not going to deny someone a tourniquet because we want to compare their rate of hemorrhage with that of someone who received a tourniquet. In those instances, identifying cause-and-effect relationships in a controlled experiment is not possible, but we can examine the relationships between variables or differences between groups. In nonexperimental research (i.e., research that is not a controlled trial), be wary if investigators use language that indicates cause-and-effect. At best in nonexperimental research, we can state that there are significant differences between groups and interventions or that there are significant correlations (associations or relationships) in measurements with a quantitative strength of the association also reported.



Summary

This discussion provided just one of many ways to discuss research basics and is not an exhaustive exhortation. Both authors approach this from a clinical perspective, with the first author from a teaching background in

educating novice clinicians and graduate students on the elements and importance of research science knowledge. We intend to continue this series by next discussing research design, methods, and essential statistical analyses.

Bibliography

Boswell C, Cannon S. *Introduction to nursing research*. Philadelphia, PA: Jones & Bartlett; 2012.

Grove SK, Burns N, Gray J. *Understanding nursing research: Building an evidence-based practice*. St. Louis, MO: Elsevier; 2014.

Laerd statistics [computer program]. Derby, UK; Lund Research. 2017. https://statistics.laerd.com/.

Polit DF, Beck CT. *Nursing research: Generating and assessing evidence for nursing practice*. New York, NY: Lippincott Williams & Wilkins; 2008.

Kemplin KR. Undergraduate and graduate research lecture collection. The University of Tennessee at Chattanooga.

"Research Wahlberg" memes available at Twitter @ ResearchMark.

Photographs of normal distribution curves and mock cholesterol concentrations from the University of Connecticut Educational Research Basics website: http://researchbasics.education.uconn.edu/averages/ and http://blogs.sas.com/content/tag/normal-distribution/.

Dr Kemplin is the Assistant Professor of Nursing Research and Statistics, The University of Tennessee at Chattanooga School of Nursing, and the director of the Research Program for the Special Operations Medical Association Scientific Assembly. E-mail: kate-kemplin@utc.edu.

SGM Bowling is the senior enlisted medical advisor of the United States Special Operations Command.

The Special Operations Medical Association's Official Journal

JSOM

JOURNAL *of* SPECIAL OPERATIONS MEDICINE™

Summer 2017
Volume 17, Edition 2

THE JOURNAL FOR OPERATIONAL MEDICINE AND TACTICAL CASUALTY CARE

Inside this Issue:

- TCCC Guidelines Change 16-03
- Assessment of Trainer Skill
- Bioelectric Dressing for Blister Management
- Rapid Vision Correction by SOF
- Role 1 Resuscitation Team and REBOA
- Preparing to Deploy to a Medically Austere Theater
- Manikin Human-Patient Simulator Training
- Complication of Attempted Surgical Airway
- Albumin Fluid Resuscitation in TCCC
- QuikClot® Combat Gauze™ Use in Afghanistan
- Ongoing Series: Clinical Corner, Human Performance Optimization, Infectious Diseases, Injury Prevention, Prolonged Field Care, SOFsono Ultrasound Series, Special Talk, Unconventional Medicine, Book Review, TCCC Updates, TacMed Updates, *and more!*

Dedicated to the Indomitable Spirit and Sacrifices of the SOF Medic

A Peer-Reviewed Journal That Brings Together the Global Interests of Special Operations' First Responders