ECRONICON
OPEN ACCESS

EC ORTHOPAEDICS
Research Article

Causation, Relationship, Association, and Correlation: Narrative Review

Mark E Murdock*

Palmer College of Chiropractic Florida, Port Orange, FL, USA

***Corresponding Author:** Mark E Murdock, Palmer College of Chiropractic Florida, Port Orange, FL, USA.

Received: February 13, 2022; **Published:** March 10, 2022

Abstract

Introduction: In this narrative review, the author explains the terms causation, relationship, association, and correlation. Causation is addressed through Hill's criterion which includes strength of association. Relationship and association are vague terms. Correlation is explained through definitions, types, examples and incorrect usages. Correlation is important to determine which treatments need to be investigated at a higher level.

Methods: Literature review was based upon convenience sample of materials used to teach the subject and Pubmed search limited to free full text.

Results: Deliverables include explanations and diagrams of Hill's criterion and types of correlation.

Discussion: Potential problems and solutions are discussed including the improper use of Pearson's r in diagnostic statistics.

Conclusion: This provides a resource for beginning to understand the complexities of causation and elucidating features of correlation.

Keywords: *Causation; Relationship; Association; Correlation; Pearson R; Evidence Based Medicine; EBM; EBCP; Education*

Introduction

Basic understanding, examples

In this article, I will explain the loose criterion for causation called Hill's Criterion and focus on the criterion called "strength of association" which includes correlation. Correlation is an effect size for how close data fits a model, direction of the model and, by extension ("R squared"), the amount of variation explained by the model. Perhaps this will stimulate more discussion to tighten the criterion for causation. First, I want to build the basic understanding of variables and how they may be related or not. Research is often utilized to determine what cause (C) causes (C) an effect (E): C -> E. We would like to use these relationships to control the effects by manipulating the causes. Perhaps we want to find causes for decreasing pain in orthopedic conditions. These causes are often termed predictors or independent variables while the effects are called outcomes or dependent variables. Unfortunately, we might not know if the causes (C) are really causes or the effects (E) can really be called effects. Hence, we will call them both variables since the causation is unknown (I took off the pre-terms independent and dependent) and their values are variable (definition of variable). These variables could be not related or related. Relationship types include (but are not limited to) many types and are depicted in table 1. Meanwhile, types of non-relationships are in table 2.

Type of relationship	Notation	Example
Linear	y=mx+b, E=mC+b	F=ma; dG=dH-TdS
Quadratic; Sinusoidal, etc.		E=mcc; y=sinx
Interaction	En + B + ChemA -> En + B + ChemB EnB + ChemA = ChemB	The necessity of an enzyme and B vitamins to convert homocysteine to methionine; both the enzyme and B vitamins must be present

Table 1: *Relationships.*

Type of Non-relationships	Notation	Example
Time dependent artificial	A2 = A1 X t B2 = B1 X t A2~B2	find any two graphs of things that go up over time and pretend they are related. More specifically, both age and price of butter go up over time; does higher butter prices cause us to age. Just google positive linear trends and make up your own positive correlations.
Chance	A1 to A20 investigated A15 is correlated to B	A researcher can investigate too many variables and accidently get a correlation with some of the variables. At the 95% confidence interval, 20 variable most likely provide about one chance correlation. However, these spurious correlations do not normally show up consistently in repeated studies unless related to time.
Mismatched		Association found is not what one really wants. E.g. If a pretest (quizzes, etc.) has all the answers then the test will be correlated yet the cause is not because the student necessarily knows the information. They would probably have more difficulty if tested with a different scenario.
Tertium quid	A ->B A -> C B~C	weather causes ice cream sales and coat wearing behavior; but, if I stop wearing my coat in the winter, will I want ice cream

Table 2: *Non-relationships.*

Methods

Literature review was based upon convenience sample of materials used to teach the subject and Pubmed search limited to free full text using theterms below. Search was filtered to meta-analysis, RCT, review, systematic review. Term results and trials included:

- 0 causation, relationship, association, correlation, “pearson r:
- 0 causation, correlation, “pearson r”
- 10 causation, “pearson r” (no relevance)
- 36,975 causation (filtered to reviews only, too many titles, found 2 related to topic on first page).

Results

Deliverables from the literature search include explanations and diagrams of Hill’s criterion, types of correlation and diagnostic statistics. Although I have used these terms rather loosely above, I will now try to define the terms causation, relationship, association, and correlation. The criterion for causation as per Hill’s criterion [1-3] is included in table 3.

Hill's Criterion	Main idea	Examples 1. Billiards, 2. Pain reduction
Strength [2][1] of association [3p96]	Meta-analysis shows large effect size (Cohen's d, odds ratio, Pearson's r etc.)	1. Cue ball struck at particular force in Newtons causes ball to move 4 feet 2. Amount of pain decreased by a medication
Consistency [2][3] with other knowledge [3]	Reproducibility in different settings	1. Cue ball moves just like many other things struck by cue stick (except large boulders) 2. The treatment seems to decrease pain in other states (Florida, Utah, etc.), socio-economic classes and facilities (hospital, clinic, etc.)
Specificity [2] [3]	When x is removed, y does not happen (aka necessary cause) [4]	1. Cue ball stays in place if not struck most of the time 2. The person does not seem to get better with anything else other than that treatment.
Temporality [2] [3p97]	Cause happens before effect	1. Struck cue ball moves; not moving cue ball caused person to strike it. Although, birthday cards do not cause the birthday to happen. 2. The pain med was given and the patient felt better when the serum levels started to rise (not immediately after swallowing)
Biological gradient or dose-response relationship [2] [3]	More x then more y, related to strength [1p5]	1. Cue ball struck by more force moves more 2. The treatment decreases more pain with more dose within the therapeutic range (side effects should be considered)
Biological [3] Plausibility [2][3]	Jives with existing mechanisms of pathophysiology	1. Studies show how round things that are struck move after a threshold is met 2. Study showing how the pain medication blocks free end nerves that relate pain
Other explanations [2]	Take out other possibilities	1. Was cue ball moved because of earthquake or someone bumping table? 2. Was the person taking heroin while they were on the pain medication and the real cause of pain decreasing was the illegal drug?
Experimental [3] confirmation [2]	RCT show strong effect, p-value not chance, CIs for Tx … Cohort study for Harm … Animal studies	1. When we plot the amount of force with how far the cue ball moves does it show a positive correlation when we take out other factors that is greater than chance? 2. When we plot pain verses time, does it seem to decrease the pain whether the pain is slight or significant or unbearable compared to no treatment and by chance?
Biological [3] Coherence [2,3]	Jives with existing theory and knowledge	1. Newton's laws of motion address this when modified for tilt of the table top, wind from AC and friction. 2. Does treatment decrease pain related to Gate Theory?
Analogy [3]	Jives with other factors and related diseases	1. Ball moves when hit by bat is analogous to cue ball hit by cue stick. 2. Other meds that use the same pathway to decrease pain (Cortex, nuclei, spinal cord synapse, PNS, or receptor).

Table 3: *Hill's criterion.*

Some questions seem to emerge from this list. Is biological plausibility, biological coherence and analogy similar criterion or overlapping? Is strength and dose response overlapping? What combinations of the criterion reach a threshold for causality? Also does this model for causality pass its own criterion (i.e. is this model able to diagnose a cause)? Does this model quantify the probability of cause and cause overlap (like a partial correlation). The answers to these questions require more consideration and shared discussion by different individuals with their unique perspectives. In addition, one must also consider the use of logical arguments (modus ponens, reductio absurdum) and avoidance of logical fallacies (appeal to force, appeal to authority, ad hominem, etc). I will now focus on the first criterion "Strength of association" which will be the focus of this article.

Strength of association is one of the criteria for causation according to Hill's criterion. Association and relationship are general terms to convey the idea that there is a connection between two or more variables. Connections between the variables could be addition, subtraction, multiplication, division, bracketing, absolute value, exponents, log, sine, derivative, integral, coefficients, coincidence, loose connections etc. Definitions of association and relationship are often vague and do not include mathematical language. Mathematical models are used in research to refine associations. Correlation is a more specific term used for relationships that are linear as is defined by different mathematical equations. Different types of correlation use different equations. Table 4 shows the main types of correlations.

Type of Correlation	Criterion for usage
Pearson's r	Parametric (normal, linear, etc.) [1][5]
Spearman's rho	Non-parametric [6p279] [1]
Kendall's tau	Non-parametric (small data sets with many tied ranks) [6p279] [1]
Point-biserial	Discrete dichotomous data (E.g. Pregnant or not) [6p271]
Biserial	continuous dichotomous data (Pass or Fail on exams, underlying continuous data of scores) [6p271]

Table 4: *Correlation.*

Idea of a model

As background for correlation, we need to understand the concept of a model. The model could be simple such as if I am paid $15/hr and I work 1 hour then I get $15. Then you add taxes into the model and it becomes more complex. We could represent the simple model as a linear relationship in the form of:

{1} Y=mx+b

{2} Amt paid = (hours I work) times (wage/hour) + 0

I put 0 at the end because if we do not work then we do not get a baseline amount of money added to our check. Therefore, we can simplify this further as:

{3} Y = mx

{4} Amt paid = (hours I work) times (wage/hour)

We could use this formula to predict what our wages will be at the end of a week. However, we would be surprised at the end of the week since we did not include all of the variables. We could complicate this as follows:

{5} Amt I can deposit in my checking account = (hours I work) times (wage/hour) + tips - social security - medicare - withholding.

Notice that I slipped "tips" into the equation. This makes a good analogy to random error. Just like we cannot control the tips and it varies, variables often have some uncontrolled variation that exists. If we search hard enough, we may find some variables that seem to account for the variation. Hopefully, we can account for the majority of the variation so we do not come short at the end of the month to pay our bills. In other words, hopefully, we can have enough control over the causes to predict the effect we want. If our prediction falls short then we can manipulate the variable (work more hours, decrease withholding, get another job, ask for a raise, etc).

Correlation

Correlation basically shows us how well our observations jive with our model. Classically, correlation is the "measure of the strength of relationship between two variables" [6p82] and is a standardized effect size [6p82, 266]. Other standardized effect sizes include Cohen's d and odds ratios that are meant to "quantify the strength of an experimental effect" [6p82, 7]. Cohen's d is better if the group sizes are different [6p83]. Correlation reveals the following about a set of data's fit to a model: 1. How close the data are to the best fit line or curve on a graph, 2. Whether the relationship is directly (+) or inversely proportional (-) [7]. Correlation reveals dependable relationships between predictor and the outcome; between stimulus and response; even two non-related entities. Scatter plots (graphs between x and y) are often used to illustrate correlation. We will use a previously studied phenomenon regarding students that use their cell phones and possible relationship to their GPAs [8]. These are some outcomes:

1. Negative correlation: increase usage ~ decreased GPA
2. Positive correlation: increase usage ~ increased GPA
3. Zero correlation: variable usage ~ GPA same OR usage same ~ variable GPA.

Figure 1 shows scatterplots of pretend data that depict the first 2 outcomes (1. and 2.). The graph on the left depicts a negative correlation. This is where those that used their cell phones more had lower GPAs. Notice I did not say that more cell phone usage caused the lower GPA because correlation does not tell you the direction of the cause and whether there really is a cause. The graph on the right depicts a positive correlation. Notice that those that used their cell phones more had higher GPAs.

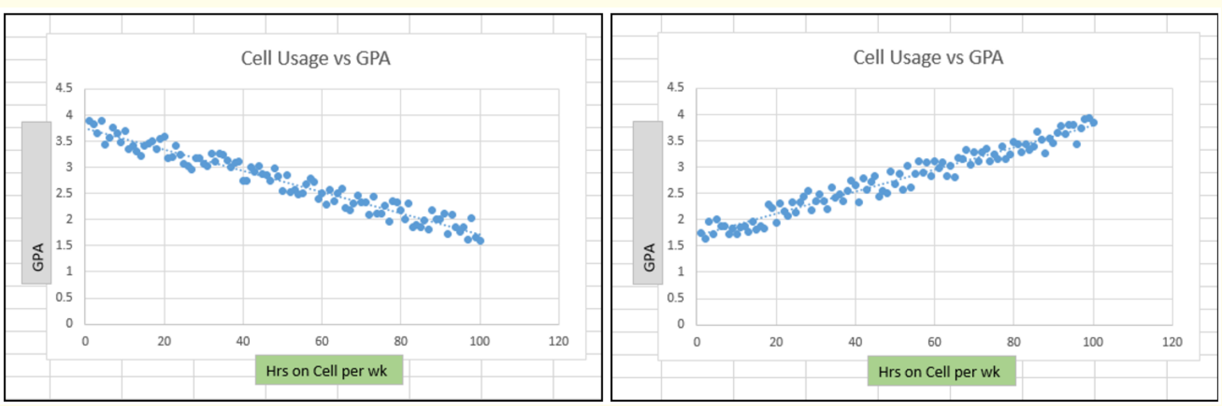

Figure 1: *Scatterplots of "make believe data" to show how the data can be in a linear fashion, a regression line can be fitted to the data, and the association determined to be positive or negative.*

We notice that the data points seem to be forming around a line. With linear regression, the line can be determined in the form of y=mx+b. Notice that the dots do not line up exactly on the line. They are scattered around the line.

There are many types of correlation: bivariate [6p267], partial [6p267] and semi-partial. Bivariate correlation includes 2 variables while partial includes 3 (1 effect, 2 proposed causes). Bivariate correlation can be parametric or non-parametric. Parametric includes Pearson's r while non-parametric includes Spearman's rho or Kendall's tau. Partial correlation will be discussed after Pearson's r.

Pearson's r (aka Pearson product moment correlation coefficient)

Pearson's r is a bivariate correlation statistic for linear relationships that is parametric. Assumptions of Pearson's r includes the following for parametric statistics: 1. Linear (if not linear then perform data transformation [6p271]), 2. Level of measurement (interval), discrete dichotomous data (E.g. Pregnant or not) use point-biserial correlation, continuous dichotomy (Pass or Fail on exams) use biserial correlation ([6p271], 3. Independent (no coregulation), 4. Normally distributed aka Bell shaped curve (+ relationship with the n number, > 30 [9], 45 is helpful due to drop outs, 100 is reasonably large [6p270], if significance/p-value and CI is needed then perform bootstrapping, if n is small then use non-parametric statistics), 5. Low heterogeneity 6. Not overfit, 7. No autocorrelation. Autocorrelations is where the values at a particular time effect the values in the future. An example of autocorrelation would be the movement of a spring (stretched and let go) or the movement of the S&P500 index fund over time. Non-parametric statistics for correlation include Spearman's rho or Kendall's Tau. Regarding Pearson's r, let us pretend that the phone usage data had all of the dots on the regression line. That would be a perfect correlation where Pearson r would be +1 if it there was a positive slope or -1 if the slope was negative [6p267]. Figure 2 depicts values where the Pearson's r is +1 or -1.

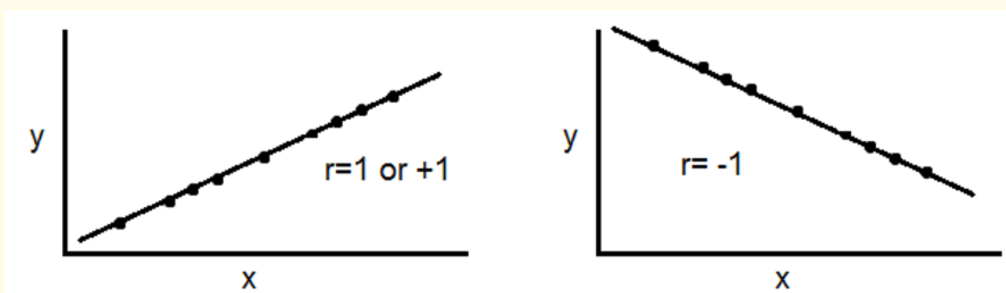

Figure 2: *Pearson's r represented on scatterplots with a value of + or – 1 since the dots representing the data are right on the line.*

On the other extreme, if the data were scattered all over without any relationship to the line, then that would be zero correlation or no correlation (r = 0). If the data points were not on the line, then what would be the value of Pearson r? Then Pearson's r would be between -1 and 0 or 0 and 1. Notice that Pearson's r is constrained to be within the inclusive range of -1 to 1 [6p82]. Figure 3 shows a plot from Microsoft Excel (Microsoft Corp) showing a regression line with a Pearson's r of 0.50. This was calculated by performing the square root of R (R squared) that is given on the graph. You can see that the slope of the regression line and the value of Pearson r are not the same (m = 1.8044, r = 0.50). Pearson r of 0.50 means that the correlation is less than perfect. This would show on a graph of that data where the dots are scattered from the line.

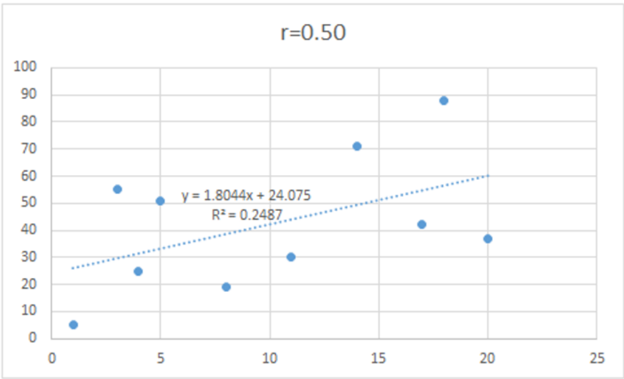

Figure 3: *Scatterplot with linear regression from Microsoft Excel including equation with slope, y-intercept, and "R squared". Taking the square root of "R squared" gives Pearson's r.*

Is 0.50 a meaningful correlation? As a rule of thumb, 0.1=small, 0.3=medium and 0.5=large. So, yes, 0.50 would be large and perhaps meaningful [6p82]. Of course this depends on how much variation you can accept in the context of your application [6p267]. Correlation with horseshoes and hand grenades has more leeway than the context of snippers. There are many examples of Pearson's r being used in the literature as follows: linear regression between quiz utilization and final grades [10], comparing ADLs and fibromyalgia surveys [11], correlation between Oswestry Disability Index (ODI) and low back pain [12] and moving Pearson's r average [13].

Other considerations of Pearson's r include determining confidence intervals [6p269] and p-values [6p268, 14]. P-values tells you whether the correlation could have happened by chance. If the p-value is greater than 0.05 then it is more likely that what was observed was by chance. P-values are dependent upon the amount of data subjects (n-number), the alpha level (normally 0.05) and the spread of the data (variance, standard deviation).

Partial correlation

Partial correlation helps to determine what percentage of the variation is correlated with which predictor variable (2 or more). Figure 4 represents a scenario where 40% (white blocks) of the pain reduction is due to unknown cause, 10% due to x1 (yellow, Eg. medication 1) 20% due to x2 (blue, Eg. medication 2) and 30% due to an interaction between med A and B. The equation would look similar to this:

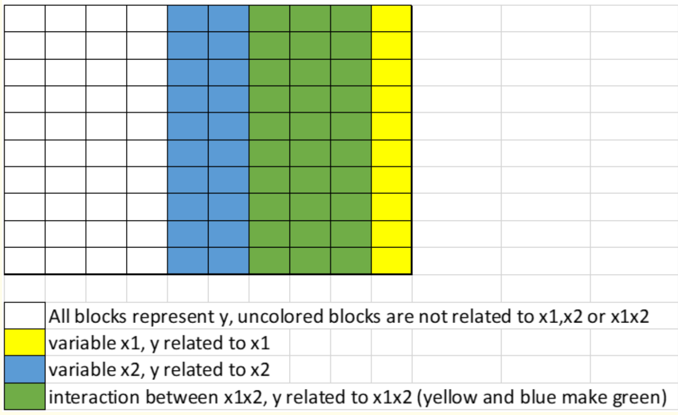

Figure 4: *Partial correlation representing x1 (yellow) related to 10% of the variation in y. x2 = 20%. However, x1 and x2 have a 30% effect by interaction. The interaction is an overlap of x1 and x2 and their effect on y.*

{8} y = m1x1 + m2x2 +m12x1x2.

Notice the coefficient m12 is the effect from an "interaction" between x1 and x2. The standardized equation would look like this if we add the correlation coefficients:

{9} y = 0.10x1 + 0.20x2 + .30x1x2 + .40x3

Where x3 is unknown causes of variation.

Discussion

Now we will discuss the implications of correlation being used in Hill's criterion for causation. I will discuss potential problems and solutions.

Incorrect usage of Pearson's r

Pearson's r is sometimes used incorrectly for diagnostic reliability: inter-examiner reliability and agreement. The diagnostic statistics figure gives the framework for this discussion (Figure 5). This figure shows that Pearson's r should not be used for inter-examiner reliability [15] or agreement [16]. Inter-examiner reliability is to determine if multiple examiners will get the same result of a test (E.g. positive/negative or 120/80). Kappa, ICCs and MAD are for inter-examiner reliability [17-20]; meanwhile, Bland Altman limits of agreement [16,21-23] and Ordinary Least Products (OLP) [24,25] are for agreement [26,27]. Agreement is whether multiple tools will agree regarding the same finding (E.g. manual or electronic blood pressure cuffs agreeing). Pearson's r has been used for validity as well [28]; although likelihood ratios are considered the best. In 1991, Haas reported many studies that used the incorrect diagnostic statistic or conclusion [15]. Studies still use Pearson's r in diagnostics [29-31]. Correlation studies the relationship between two variables; not the difference [32]. Griffiths gave an example of a BP measuring device that is always 15mmHg higher than it should be compared to the standard [33]. There is a large part that was cut off the end of this paper including: Pearson's confusion, problems with correlation, correlation to causation, conclusions, funding, disclosures, etc. Please see attached document.

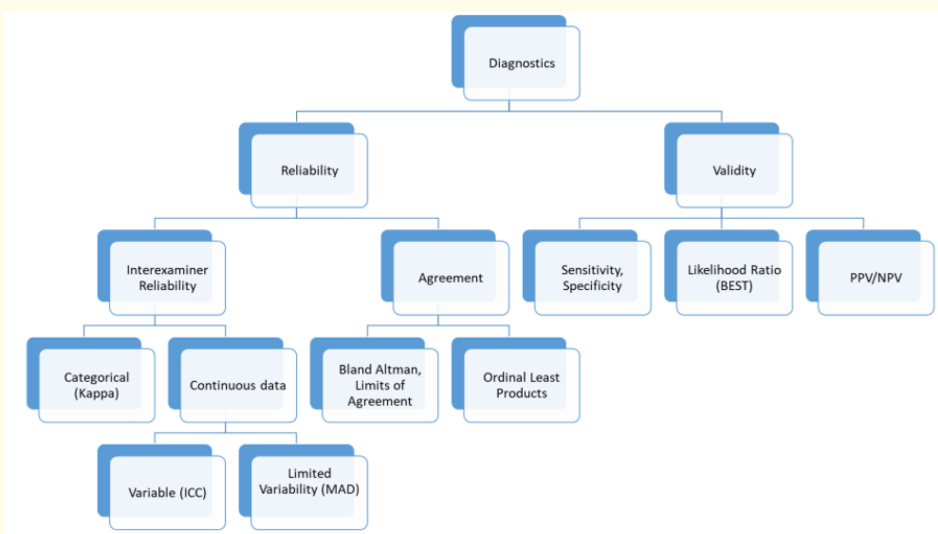

Figure 5: *Diagnostic statistics.*

Pearson's confusion

The statistician Carl Pearson has his name associated with many named statistical tests and this may cause some confusion. These tests include: Pearson's Chi square [34] and Clopper Pearson's method [35].

Problems with correlation

Problems with correlation include: 1. Does not tell the direction of causality [6p270], 2. Does not show causation (Tertium quid [6p270]), 3. Related yet does not give you the slope of the line, 4. Limited to assumptions.

Correlation to causation

Once a correlation is found, it might be prudent to perform multiple randomized controlled trials (RCTs) to test the possible relationship to get closer to causation. With the RCT, one can work on some of the other Hill's criterion. When RCTs are performed they often will use difference of means tests (DOMTs) which are grouped as parametric or non-parametric. Parametric tests include t-test, ANOVA, MANOVA. Non-Parametric tests includes Wilcoxon, Freidman's and Kruskal Wallis. These tests will give the slope of the line (m) and a p-value for the slope. Often the output from DOMT statistics programs will give the Pearson's r and p-value for the Pearson's r as well. The slope and Pearson's r are associated in this equation:

{6} m * sdx/sdy = r

{7} rise(y)/sdy * run(x)/sdx = r.

In this equation m = slope, sdy = standard deviation of y, sdx = standard deviation of x, r = Pearson's r, * = times, multiply. If the standard deviation of x and y remain constant then an increase in m will be reflected as an increase in r. Thus, r is an effect size. It is also standardized since we are dividing the signal (x) by the noice (standard deviation); and the same with y. So we can call it a standardized effect size. If we use the equation to predict the effect (y) then it does not take into account the noise (standard deviation) nor the effect of units. Pearson's r takes into account the noise and the units are divided away. If the rise is 4 meters divided by the sdy = 2 meters then the rise/sdy is 2 (notice that meters divided by meters = 1 just like 4/4 = 1). So Pearson's r is unit-less and can be compared with other contexts. We may say that study 1 shows a Pearson's r of 0.50 while another is 0.30. We can say that study 1 showed more of an effect.

Another thought is if the standard deviation of x and y are the same (no matter how high or low) then sdx/sdy will be close to 1 and the slope will be the same as the Pearson's r. This seems rather concerning to a person that does not like variation; however, this will affect the p-value of the "r squared". The predictability of the y value in the model can be reflected by the sum of squares error. If we quantify the differences between the model and the data ("real world") and call that the error, we can square that sum of all the error to minimize close values and amplify far values. We can show that a better model will have a decrease in the sum of squares error. The problem with this is that a model can be overfitted and account for variations that are caused by one entity with another.

So how do we get from something having correlation to causation? One thought might be that if it is highly correlated in many studies (such as in a meta-analysis that has many high quality studies included like RCTs) and in seems to reflect real life then we would only start questioning the relationship when we see aberrant outcomes. If you used the fertilizer blend that correlates to a big harvest in the research and it worked great in your garden then you can at least pretend you have causation. Why did I say pretend? It might be that you just have good plants that can extract what they need from the soil. However, if you have some bad harvests then the causation is questioned and more variables should be included in the equation: y=m1x1 + m2m2. Maybe your neighbor started spraying weed killer on the fence line (m2 = the amount of weed killer) or it has not rained much lately, or a microbe has infected the plants? Thus, the model becomes better with time and becomes predictive enough to control the outcomes. Nevertheless, criterion for causation have been discussed in other papers at length [4] and has caused the questioning of long held beliefs [36]; or perhaps not since there is not a universally defensible criterion for the term caused.

Conclusion

In this narrative review, I explained the terms causation, relationship, association, and correlation. Causation was addressed through exploring Hill's criterion which includes strength of association. Relationship and association are vague terms that include correlation. Correlation was explained through definitions, types, examples and incorrect usages. Correlation does not mean causation; however, it is part of Hill's criterion for causation called "strength of association". Hill's criterion does not have a mathematical threshold than can be used to compare to chance. Granted experiments can make comparisons to chance through p-values. However, I am not aware of any thresholds for causation. The main point is that there is not an established criterion that can be mathematically determined to reach a threshold of causality. This provides a resource for beginning to understand the complexities of causation and elucidating features of correlation.

Funding Support

No funding was required for this research.

Other Disclosures

The author reports no conflict of interest. The author is solely responsible for the content and writing of this original article.

Ethical Approval

This article does not constitute human subjects research.

Bibliography

1. Mirtz TA., *et al*. "An epidemiological examination of the subluxation construct using Hill's criteria of causation". *Chiropractic and Osteopathy* 17 (2009): 13.
2. Tuchin P. "Chiropractic and Stroke: Association or Causation?". *International Journal of Clinical Practice* 67.9 (2013): 825-833.
3. Moye Lemuel A. "Statistical Reasoning in Medicine: The Intuitive P-Value Primer". 2nd Edition, Springer (2006).
4. Parascandola M and Weed DL. "Causation in epidemiology". *Journal of Epidemiology and Community Health* 55.12 (2001): 905-912.
5. Miles J and Banyard P. "Understanding and Using Statistics in Psychology". SAGE Publications: Los Angeles (2007).
6. Field Andy. "Discovering Statistics Using IBM SPSS Statistics?" 4th Edition, Sage, Los Angeles, Book (2013).
7. Haneline MT. "Evidence-Based Chiropractic Practice". Jones and Bartlett Publishers, Boston (2007): 136.
8. Lepp A., *et al*. "The Relationship Between Cell Phone Use and Academic Performance in a Sample of U.S. College Students". SAGE Open (2015): 1-9.
9. Triola. "Elementary Statistics". 8th Edition, Textbook, Pearson: Boston (2001).

10. Choi M and Murdock ME. "The Effect of Online Formative Self-Assessment on Academic Performance of Chiropractic Students". *World Journal of Educational Research* 7.1 (2020): 52.
11. Collado-Mateo D., *et al*. "Stair negotiation in women with fibromyalgia". *Medicine* 96.43 (2017): e8364.
12. Cambron JA., *et al*. "Shoe Orthotics for the Treatment of Chronic Low Back Pain: A Randomized Controlled Trial". *Archives of Physical Medicine and Rehabilitation* 98.9 (2017): 1752-1762.
13. McCoy M., *et al*. "Intra-Examiner and Inter-Examiner Reproducibility of Paraspinal Thermography". *Plos One* 6.2 (2011): e16535.
14. Biau DJ., *et al*. "P Value and the Theory of Hypothesis Testing". *Clinical Orthopaedics and Related Research* 468.3 (2010): 885-892.
15. Haas M. "The Reliability of Reliability". *Journal of Manipulative and Physiological Therapeutics* 14.3 (1991): 199-208.
16. Bland JM and Altman DG. "Applying the right statistics: analyses of measurement studies". *Ultrasound in Obstetrics and Gynecology* 22.1 (2003): 85-93.
17. Koo TK and Li MY. "A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research". *Journal of Chiropractic Medicine* 15.2 (2016): 155-163.
18. Landers R. "Computing Intraclass Correlations (ICC) as Estimates of Interrater Reliability in SPSS". The Winnower (2015): 1-4.
19. Aartun E., *et al*. "Screening of the spine in adolescents: inter- and intra-rater reliability and measurement error of commonly used clinical tests". *BMC Musculoskeletal Disorders* 15 (2014): 37.
20. Whiting P., *et al*. "Development and validation of methods for assessing the quality of diagnostic accuracy studies". *Health Technology Assessment* 8.25 (2004): 1-234.
21. Allagaert K., *et al*. "Tympanic, Infrared Skin, and Temporal Artery Scan Thermometers Compared with Rectal Measurement in Children: A Real-Life Assessment". *Current Therapeutic Research* 76 (2014): 34-38.
22. Beckstead JW. "Agreement, reliability, and bias in measurement: Commentary on Bland and Altman (1986; 2020)". *International Journal of Nursing Studies* 48.1 (2011): 134-135.
23. Liu J., *et al*. "Correlation and agreement: overview and clarification of competing concepts and measures". *Shanghai Archives of Psychiatry* 28.2 (2016): 115-120.
24. Ludbrook J. "Statistical techniques for comparing measures and methods of measurement: a critical review". *Clinical and Experimental Pharmacology and Physiology* 29.7 (2002): 527-536.
25. Woodman RJ. "Ordinary Least Products: Let's welcome back an old friend". *Clinical and Experimental Pharmacology and Physiology* 39 (2012): 325-326.
26. Watson PF and Petrie A. "Method agreement analysis: A review of correct methodology". *Theriogenology* 73.9 (2010): 1167-1179.
27. Zaki R., *et al*. "Statistical Methods Used to Test for Agreement of Medical Instruments Measuring Continuous Variables Method Comparison Studies: A Systematic Review". *Plos One* 7.5 (2012): e37908.
28. Alghadir AH., *et al*. "Test-retest reliability, validity, and minimum detectable change of visual analog, numerical rating, and verbal rating scales for measurement of osteoarthritic knee pain". *Journal of Pain Research* 11 (2018): 851-856.

29. Cracknell J., *et al*. "Intra- and inter-observer reliability of the Cobb measurement by chiropractic interns using digital evaluation methods". *Journal of the Canadian Chiropractic Association* 59.3 (2015): 261-268.

30. Ferriera-Valente MA., *et al*. "Validity of four pain intensity rating scales". *Pain* 152.10 (2011): 2399-2404.

31. Triano JJ., *et al*. "Review of methods used by chiropractors to determine the site for applying manipulation". *Chiropractic and Manual Therapies* 21.1 (2013): 36.

32. Giavarina D. "Understanding Bland Altman analysis". *Biochemia Medica* 25.2 (2015): 141-151.

33. Griffiths PD and Murrells T. "Reliability assessment and approaches to determining agreement between measurements: Classic methods paper". *International Journal of Nursing Studies* 47.8 (2010): 937-938.

34. Hawk C., *et al*. "Implementation of a Course on Wellness Concepts into a Chiropractic College Curriculum". *Journal of Manipulative and Physiological Therapeutics* 28.6 (2005): 423-428.

35. Rabin A., *et al*. "The Sensitivity of the Seated Straight-Leg Raise Test Compared with the Supine Straight-Leg Raise in Patients Presenting with Magnetic Resonance Imaging Evidence of Lumbar Nerve Root Compression". *Archives of Physical Medicine and Rehabilitation* 88.7 (2007): 840-843.

36. Baedke J. "What's Wrong with Evolutionary Causation?". *Acta Biotheoretica* 69.1 (2021): 79-89.

Volume 13 Issue 4 April 2022