



Published in final edited form as:

Postgrad Med. 2017 November ; 129(8): 902–906. doi:10.1080/00325481.2017.1383820.

Interpretation of correlations in clinical research

Man Hung¹, Jerry Bounsanga¹, and Maren Wright Voss¹

¹University of Utah Department of Orthopaedic Surgery Operations, Salt Lake City, Utah, USA

Abstract

Background: Critically analyzing research is a key skill in evidence-based practice and requires knowledge of research methods, results interpretation, and applications, all of which rely on a foundation based in statistics. Evidence-based practice makes high demands on trained medical professionals to interpret an ever-expanding array of research evidence.

Objective: As clinical training emphasizes medical care rather than statistics, it is useful to review the basics of statistical methods and what they mean for interpreting clinical studies.

Methods: We reviewed the basic concepts of correlational associations, violations of normality, unobserved variable bias, sample size, and alpha inflation. The foundations of causal inference were discussed and sound statistical analyses were examined. We discuss four ways in which correlational analysis is misused, including causal inference overreach, over-reliance on significance, alpha inflation, and sample size bias.

Results: Recent published studies in the medical field provide evidence of causal assertion overreach drawn from correlational findings. The findings present a primer on the assumptions and nature of correlational methods of analysis and urge clinicians to exercise appropriate caution as they critically analyze the evidence before them and evaluate evidence that supports practice.

Conclusion: Critically analyzing new evidence requires statistical knowledge in addition to clinical knowledge. Studies can overstate relationships, expressing causal assertions when only correlational evidence is available. Failure to account for the effect of sample size in the analyses tends to overstate the importance of predictive variables. It is important not to overemphasize the statistical significance without consideration of effect size and whether differences could be considered clinically meaningful.

Keywords

Correlation; Causal Relationships; Clinical Education; Instructional Materials; Practice-Based Research; Non-Experimental Methods; Commentary

Correspondence: Dr. M. Hung, Associate Professor, Department of Orthopaedics, University of Utah, 590 Wakara Way, Salt Lake City, UT 84108, Phone: 801-587-5372, Man.Hung@hsc.utah.edu.

Declaration of Interests

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Introduction

It is common for physicians to underestimate the relevance and importance of their training in statistics and probability, at least until its relevance becomes clear in later clinical practice.¹ Evidence based practice is the standard of care, yet evaluating the quality of evidence can be a difficult process. The British Medical Journal established clear statistical guidelines for contributions to medical journals in the 1980's.² A survey of medical residents showed near perfect agreement (95%) that understanding statistics they encounter in medical journals is important, but 75% said they lacked knowledge.³ The average score on the basic biostatistics exam for the residents surveyed was 41% correct, showing both the objective and subjective need for a stronger statistical training foundation. An international survey of practicing doctors similarly found that doctors averaged scores of 40% when tested on the basics of statistical methods and epidemiology.⁴

The efforts to improve the methods behind clinical science are ongoing. Consolidated Standards of Reporting Trials (CONSORT) have been established,⁵ along with multiple other research and statistical method guidelines in recent years.⁶ These guidelines are updated periodically⁷ and endorsed or extended by specific clinical medical groups,^{8,9} or medical journals.^{10–13} Key to many of these guidelines are attempts to ensure that bias and error are minimized in research, ensuring interpretations are meaningful and accurate. It would not be practical to detail an exhaustive list of guidance and recommendations. However, with new information and research approaches constantly coming forward, physicians must have the ability to critically evaluate the quality of evidence presented. Critically analyzing research is a key skill in evidence based practice and requires knowledge of methods, results interpretation, and applicability—all three of which require an understanding of basic statistics.¹⁴

The intention of this article is to serve as a basic primer regarding critical statistical concepts that appear in medical literature with a focus on the concept of correlation and how it is best utilized in clinical interpretation for understanding the relationships between health factors. At its foundation, correlational analysis quantifies the direction and strength of relationships between two variables. Understanding correlations can form the basis for interpreting applications of clinical research.

1. What Does Correlation Tell Us?

Correlation is concerned with association; it can look at any two measured concepts and compare their relationships. These measured concepts are often referred to as variables and are assigned letter labels (X, Y). Thus, the correlation is the measure of the relationship between X and Y, and it ranges from -1 to 1 . Its value (or coefficient) is scaled within this range to assist in interpretation, with 0 indicating no relationship between variables X and Y, and -1 or 1 indicating the ability to perfectly predict X from Y or Y from X (see Figure 1). A correlation coefficient provides two pieces of information. First, it predicts where X (the measured value of interest) falls on a line given a known value of Y. Second, it expresses a reduction in variability associated with knowing Y, telling us something about the expected range of the X value.¹⁵ Correlation takes into account the full range of scores, but as a statistical tool it is not very sensitive to scores on the very high or very low ends of our X

value. The common Pearson correlation is best used to describe linear relationships. If the pattern of association between the two variables is, for example, a “U” shaped curve, the correlation results might be low, even though a defined relationship exists.¹⁶

Correlation is not easily impacted by skewed or off-center data results. The nature of the data can be described by its parameters, like measures of central tendency, which inform us about the distribution of the data. Parametric tests make assumptions, such as the data are normally distributed, while non-parametric tests are called distribution-free tests because they make no assumptions about the distribution of the data. Yet even if the pattern of scores are not in a normal bell shaped curve or they don't create a direct linear relationship, correlations can still be reliable. A number of correlation measures have been developed to handle different types of data (non-parametric tests like the kendall rank, spearman rank correlation, phi correlation, biserial correlation, point-biserial correlation and gamma correlation). Yet even the simple Pearson correlation handles extreme violations of normality (no bell shape to the pattern of scores) and scale. The Pearson correlation was tested by randomly drawing 5,000 small samples (n=5 to n=15) from a population of 10,000 to calculate the distribution of r values yielded (small samples might challenge parametric assumptions), and was still found to be a reliable indicator of the relationship between variables.¹⁷ It is also possible to use transformations to normalize the distribution of data. Correlation can be a robust measure, in part from its ability to tolerate these violations of normally distributed data while staying sensitive to the individual case. The properties of correlation make the technique useful in interpreting the meaning derived from clinical data.

2. Use of Correlation

There are many concerns with the statistical techniques that are commonly utilized in the literature. Some concerns arise from a misunderstanding of the statistical measure and others from its misapplication. We discuss four ways in which correlational analysis is misused, including causal inference overreach, over-reliance on significance, alpha inflation, and sample size bias. Importantly, correlation is a measure of association, which is insufficient to infer causation. Correlation can only measure whether a relationship exists between two variables, but it does not indicate causal relationship. There must be a convincing body of evidence to take the next step on the path to inferring that one variable *causes* the other. Randomized controlled trials or more advanced statistical methods such as path analysis and structural equation modeling, coupled with proper research design, are needed in order to take the next step of inference in the causal chain, testing a causal hypothesis. While correlational analyses are by definition from non-experimental research, research without carefully controlled experimental conditions, it is nevertheless relevant to evidence-based practice.¹⁸ Observational methods of study can be conducted using either a cross-sectional design (a snap-shot of prevalence at one time-point), a retrospective design (looking back to compare current with past attributes) or prospective design (documenting current occurrences and following up at a future time-point to make comparisons). Correlations drawn from cross-sectional studies cannot establish the temporal relationship that links cause with effect, yet adding a retrospective or prospective observational design provides additional strength to the association and helps support hypothesis generation to then later test the causal assertions with a different research design.¹⁹

When non-experimental methods are used, it means the relationship seen between the two variables is vulnerable to bias from anything that was not measured (unobserved variables). Whether studying pain or function or treatment response, there are a host of possible factors that might be important to the observed correlational relationship between X and Y. Any given study has only measured and reported on a fraction of the potential variables of impact. These unobserved variables could potentially explain the observed relationship, so it would be premature to assume a treatment effect based on correlational data. The unobserved variables might be affecting the study variables, changing the relationship in a way that might alter the interpretation of the data. Thus interpretation of correlational findings must be quite cautious until further research is completed.

Another concern is the use (or abuse) of the term ‘statistically significant’ in correlational analysis. This concern is not new. The abuse of significance testing was noted in a 1987 review published in the *New England Journal of Medicine*. That article found that a number of components in clinical trials, such as having several measures of the outcome (i.e., multiple tests of function, health, or pain), repeated measures over time, including subgroup analyses, or multiple treatments in the same trial, can lead to a bias in reporting which exaggerates the size or importance of observed differences.²⁰ It is natural for researchers to want to thoroughly evaluate the potential difference between treatments conditions. This has been sometimes referred to as the kitchen sink approach and it presents a problem for using significance tests. Significance as a statistical procedure addresses the question of the probability of the hypothesized occurrence. If the probability (p) is less than say 5% or 1%, the researcher might feel comfortable making the assumption that the observed event was not due to chance. The 0.05 significance value was originally proposed by Sir Ronald Fisher in 1925, but the 0.05 value was never intended as a hard and fast rule.²¹ If researchers use a cut-off of $p=0.05$ to determine whether the effect they see in their research has occurred randomly by chance, running multiple tests can quickly move the needle from a rare to an expected event. Every analysis run with a $p=0.05$ alpha criteria yields a 5% chance that the “significant” finding is actually a chance occurrence, called a Type 1 error.

So we can see the difficulty that occurs when running 20 different analyses on the same data. This would produce a 64% chance that a significant p -value will show up erroneously, when there is no systematic relationship, and it is really just a chance occurrence (see Figure 2). This over analysis of the data can possibly create an interpretation that a test or treatment should be used when there was no actual treatment effect. This is the problem of alpha-inflation and it needs to be carefully considered both in conducting and interpreting correlational research. It can be corrected by planning ahead for the analyses that will be run and keeping them limited to key theoretical questions.²⁰ Additionally, alpha can be adjusted in the statistical calculation, for example with the Bonferroni correction, a procedure used to reduce alpha and statistically correct for the inflation created by multiple comparisons. It is performed by simply dividing the alpha value (α) by the number of hypotheses or measures (m) tested. If the study wanted to evaluate 5 different surgical placements, using the Bonferroni correction would adjust an original $\alpha=0.05$ by applying the formula, α/m or $0.05/5$, yielding $\alpha=0.01$, a stricter standard before a study finding would be considered significant. A Bonferroni correction is a conservative correction for multiple comparisons that reduces the Type 1 error,²² though less conservative alternatives exist like the Tukey or

the Holm-Sidak corrections. The main idea is that clinicians acknowledge the problem of multiple comparisons made in a single study and address the concern so that spurious relationships are not erroneously reported as significant. A lack of awareness about this issue can lead to naïve interpretations of study findings.

Exaggeration of significance testing leads to a third point - are the findings clinically meaningful? A significant finding does not infer a meaningful finding. This is because factors other than variance in scores influence the p -value or significance in a correlational analysis. Sample size is an important element in whether a non-random effect will be found. Small sample sizes might produce unstable, but significant, correlation estimates, so sample sizes greater than 150 to 200 have been recommended.²³ Yet, it is not uncommon for published papers to report significant effects through correlational analysis of sample sizes of less than 150 patients.^{24–26} While reporting and publishing both the significant and non-significant results are important, given the instability that comes from a small sample size, there should be caution taken with interpretation until replication studies can verify the findings.

Likewise, large samples can also be problematic. A large sample might reveal a statistically significant difference between groups, but its effect might be minimal. In a classic example, a sample of 22,000 subjects showed a highly significant ($p < .00001$) reduction in myocardial infarctions that prompted a general recommendation to take aspirin for myocardial infarction prevention.²⁷ The effect size, however, was less than a 1% reduction in risk, such that the risk of taking aspirin exceeded the benefit. Effect size is the standardized mean difference between groups and is a measure of the magnitude of between group differences. A significant p -value indicates only that a difference exists with no indication of size of the effect. Additionally, a confidence interval (CI) can be constructed for the effect size. CIs present a lower and upper range where the true population value is most likely to lie.²⁸ If a zero value is not included within the CI of the effect size, we have added assurance that the effect exists, with the size of the CI helpful in estimating the size of the effect.²⁹ It has been recommended that effect sizes or confidence intervals be included in all reported medical research so that the clinical significance of findings can be assessed.^{20,28,30}

3. Proper Interpretation of Correlation

Correlational analyses have been reported as one of the most common analytic techniques in research at the beginning of the 21st century, particularly in health and epidemiological research.¹⁵ Thus effective and proper interpretation is critical to understanding the literature. Cautious interpretation is particularly important, not only due to the interpretive concerns just detailed (causal inference overreach, over-reliance on significance, alpha inflation, and sample size bias), but also given the publication bias of journals to accept and publish studies with positive findings.³¹ If clinicians are less likely to be exposed to under-published contradictory reports, based on null findings that treatments actually had no effect, the interpretation of the positive results must necessarily be cautious until confirmed through strong evidence.

One recent clinical example of correlational findings is an inference that because Cobb angle and sagittal balance are related to symptom severity in back pain, treatments should be

aimed to improve sagittal balance. Studies used to draw these conclusions were making an important first step in identifying potentially relationships, but were not conclusive as they did not establish causal relationships, did not report effect sizes, and did not include control groups in the analyses.^{32–34} The Pearson correlation or the Spearman correlation are tools that predict X from Y or Y from X. The nature of the correlation is symmetric, so that if the variables are inferred from a reversed direction (pain predicting spine function rather than spine function predicting pain), the same prediction holds true.¹⁵ If one is looking for cause and effect, the correlational statistics cannot help. The mathematics of correlation tells us that Y is just as likely to precede X as to come after, because the prediction is the same regardless of which variable is inputted first. Effects cannot be determined directly through correlational analysis and perhaps the reverse relationship is the true relationship.

Because of the possibility of a bidirectional relationship, causal inference will be premature if relying purely on correlational statistics, no matter how many studies report the correlational finding. Correlation can be interpreted as the association between two variables. It cannot be used to indicate causal relationship. In fact, statistical tests cannot prove causal relationships but can only be used to test causal hypotheses. Misinterpretation of correlation is generally related to a lack of understanding of what a statistical test can or cannot do, as well as lacking knowledge in proper research design. Rather than jumping to an assumption of causality, the correlations should prompt the next stage of clinical research through randomly controlled clinical trials or the application of more complex statistical methods such as causal and path analysis. Perhaps a part of the tendency to jump too quickly to causal assertion arises from the nature of the questions asked in clinical research and the desire to quickly move to enhance patient care. New frameworks are emerging in the health sciences that challenge the appeal to a single cause by considering potential outcomes in a more complex ways.³⁵ Until then, understanding the nature of correlational analysis allows clinicians to be more cautious in interpreting study results.

Conclusion

Advances in research have led to many significant findings that are shaping how we diagnose and treat patients. As these findings might guide surgeons and clinicians into new treatment directions, it is important to consider the strength and nature of the research. Critically analyzing new evidence requires understanding of research methods and relevant statistical applications, all of which require an understanding of the analytic methodologies that lie behind the study findings.¹⁴ Evidence based practice is demanding new skills of trained medical professionals as they are presented with an ever-expanding array of research evidence. This short primer on the assumptions and nature of correlational methods of analysis can assist emerging physicians in understanding and exercising the appropriate caution as they critically analyze the evidence before them.

Acknowledgments

Funding

This investigation was supported by the University of Utah Department of Orthopaedics Quality Outcomes Research and Assessment, Study Design and Biostatistics Center, with funding in part from the National Center for

Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant 5UL1TR001067–02.

References

1. Miles S, Price GM, Swift L, Shepstone L, Leinster SJ. Statistics teaching in medical school: opinions of practising doctors. *BMC medical education*. 2010;10:75. [PubMed: 21050444]
2. Altman DG, Gore SM, Gardner MJ, Pocock SJ. Statistical guidelines for contributors to medical journals. *Br Med J (Clin Res Ed)*. 1983;286(6376):1489–1493.
3. Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. *JAMA*. 2007;298(9):1010–1022. [PubMed: 17785646]
4. Novack L, Jotkowitz A, Knyazer B, Novack V. Evidence-based medicine: assessment of knowledge of basic epidemiological and research methods among medical doctors. *Postgrad Med J*. 2006;82(974):817–822. [PubMed: 17148706]
5. Moher D, Schulz KF, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA*. 2001;285(15):1987–1991.
6. Vandembroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol*. 2009;62(6):594–596. [PubMed: 19181482]
7. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18. [PubMed: 20334633]
8. Matz PG, Meagher RJ, Lamer T, et al. Guideline summary review: an evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spondylolisthesis. *The spine journal : official journal of the North American Spine Society*. 2015.
9. Manchikanti L. Evidence-based medicine, systematic reviews, and guidelines in interventional pain management, part I: introduction and general considerations. *Pain physician*. 2008;11(2):161–186. [PubMed: 18354710]
10. Freedland KE, Babyak MA, McMahan RJ, Jennings JR, Golden RN, Sheps DS. Statistical guidelines for psychosomatic medicine. *Psychosom Med*. 2005;67(2):167.
11. Wilkinson L. Statistical methods in psychology journals: Guidelines and explanations. *Am Psychol*. 1999;54(8):594.
12. Armstrong RA, Davies LN, Dunne MC, Gilmartin B. Statistical guidelines for clinical studies of human vision. *Ophthalmic Physiol Opt*. 2011;31(2):123–136. [PubMed: 21309799]
13. Bond GR, Mintz J, McHugo GJ. Statistical guidelines for the Archives of PM&R. New Hampshire-Dartmouth Psychiatric Research Center. *Arch Phys Med Rehabil*. 1995;76(8):784–787.
14. Morris RW. Does EBM offer the best opportunity yet for teaching medical statistics? *Stat Med*. 2002;21(7):969–977; discussion 979–981, 983–984. [PubMed: 11921003]
15. Kraemer HC. Correlation coefficients in medical research: from product moment correlation to the odds ratio. *Stat Methods Med Res*. 2006;15(6):525–545. [PubMed: 17260922]
16. Kowalski CJ. On the effects of non-normality on the distribution of the sample product-moment correlation coefficient. *Applied Statistics*. 1972:1–12.
17. Havlicek LL, Peterson NL. Robustness of the Pearson correlation against violations of assumptions. *Perceptual and Motor Skills*. 1976;43(3f):1319–1334.
18. Thompson B, Diamond KE, McWilliam R, Snyder P, Snyder SW. Evaluating the quality of evidence from correlational research for evidence-based practice. *Except Child*. 2005;71(2):181–194.
19. Song JW, Chung KC. Observational Studies: Cohort and Case-Control Studies. *Plast Reconstr Surg*. 2010;126(6):2234–2242. [PubMed: 20697313]
20. Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med*. 1987;317(7):426–432. [PubMed: 3614286]
21. Fisher R. *Statistical Methods for Research Workers*. Oliver and Boyd Ltd, Edinburgh 1925:47.
22. Miller RG. *Simultaneous statistical inference*. New York: Springer; 1966.

23. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. *Psychol Bull.* 1988;103(2):265–275. [PubMed: 3363047]
24. Deviren V, Berven S, Kleinstueck F, Antinnes J, Smith JA, Hu SS. Predictors of flexibility and pain patterns in thoracolumbar and lumbar idiopathic scoliosis. *Spine (Phila Pa 1976).* 2002;27(21):2346–2349. [PubMed: 12438982]
25. Ploumis A, Liu H, Mehbod AA, Transfeldt EE, Winter RB. A correlation of radiographic and functional measurements in adult degenerative scoliosis. *Spine (Phila Pa 1976).* 2009;34(15):1581–1584. [PubMed: 19564768]
26. Schwab FJ, Smith VA, Biserni M, Gamez L, Farcy JP, Pagala M. Adult scoliosis: a quantitative radiographic and clinical analysis. *Spine (Phila Pa 1976).* 2002;27(4):387–392. [PubMed: 11840105]
27. Bartolucci AA, Tendra M, Howard G. Meta-analysis of multiple primary prevention trials of cardiovascular events using aspirin. *Am J Cardiol.* 2011;107(12):1796–1801. [PubMed: 21481826]
28. Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed).* 1986;292(6522):746–750.
29. Kelley K, Preacher KJ. On effect size. *Psychol Methods.* 2012;17(2):137. [PubMed: 22545595]
30. Sullivan GM, Feinn R. Using Effect Size—or Why the P Value Is Not Enough. *J Grad Med Educ.* 2012;4(3):279–282. [PubMed: 23997866]
31. Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *The Cochrane database of systematic reviews.* 2009(1):Mr000006. [PubMed: 19160345]
32. Glassman SD, Bridwell K, Dimar JR, Horton W, Berven S, Schwab F. The impact of positive sagittal balance in adult spinal deformity. *Spine (Phila Pa 1976).* 2005;30(18):2024–2029. [PubMed: 16166889]
33. Glassman SD, Berven S, Bridwell K, Horton W, Dimar JR. Correlation of radiographic parameters and clinical symptoms in adult scoliosis. *Spine (Phila Pa 1976).* 2005;30(6):682–688. [PubMed: 15770185]
34. Johnson RD, Valore A, Villaminar A, Comisso M, Balsano M. Sagittal balance and pelvic parameters—a paradigm shift in spinal surgery. *J Clin Neurosci.* 2013;20(2):191–196. [PubMed: 23232102]
35. Glass TA, Goodman SN, Hernan MA, Samet JM. Causal inference in public health. *Annu Rev Public Health.* 2013;34:61–75. [PubMed: 23297653]

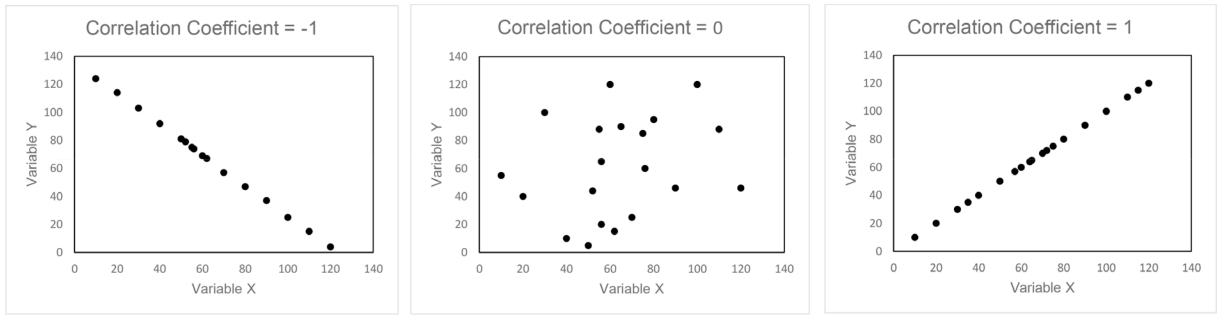


Figure 1:

1) Perfect negative correlation between two variables; 2) No patterned relationship between 2 variables; 3) Perfect positive correlation between two variables

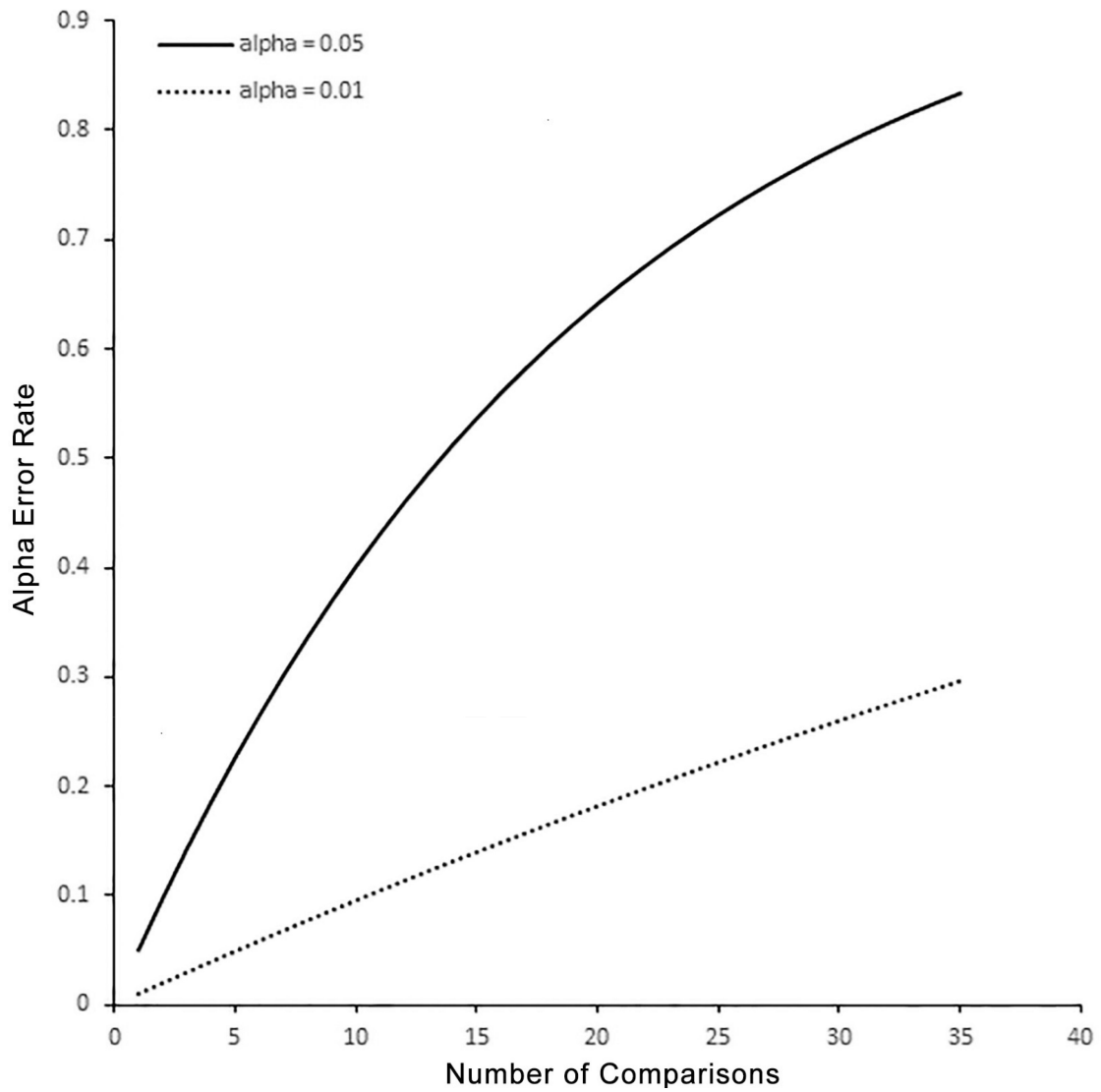


Figure 2:

As the number of comparisons increases, the alpha error rate increases; running 20 comparisons with no correction factor and a significance level set at 0.05 will result in 64% chance of a relationship appearing to be significant (better than 50/50 odds).