Review

Determining a Meaningful R-squared Value in Clinical Medicine

Avi Gupta[1], Thor S. Stead, ScB[2], Latha Ganti, MD, MS, MBA[2,3]

[1] Seminole Public Schools, [2] Brown University, [3] Orlando College of Osteopathic Medicine

Academic Medicine & Surgery

Introduction

In clinical medicine, determining a meaningful R^2 value requires careful consideration of various factors. The R-squared (R^2) value is a statistical measure used to assess the extent to which independent variables explain the dependent variable in regression models. Its interpretation in clinical medicine is very context-dependent and lacks a definitive threshold.

While a higher R^2 suggests a stronger relationship between variables, smaller R-squared values will still hold relevance, especially for multifactorial clinical outcomes. Due to clinical medicine's genetic, environmental, and behavioral factors, high R^2 values may not be realistic or expected in this context.

Methods

This study is a narrative review encompassing expert opinions, results from randomized controlled trials (RCTs), and observational studies relating to the use and interpretation of the coefficient of variance (R^2) in clinical medicine.

Results

30 studies spanning literature across various disciplines within clinical medicine were evaluated and synthesized to provide a contextualized, nuanced approach to interpreting the R^2 in medical literature.
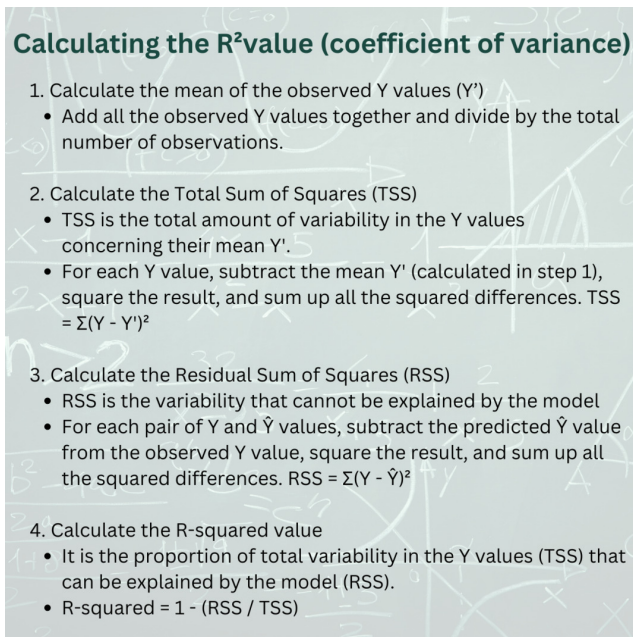
Conclusions

A nuanced understanding of the significance of R^2 values in clinical medicine requires considering the research question, clinical context, complexity of the phenomenon, and contextual comparisons. Integrating statistical rigor and clinical judgment helps derive meaningful insights from clinical data. This paper highlights key considerations for evaluating a "good" R^2 value in clinical medicine and concludes that an R^2 of >15% is a generally a meaningful value in clinical research.

INTRODUCTION

The coefficient of variance (R^2 value) is a measure used in statistics that represents the amount of variance in the outcome which can be explained by the independent variable(s). It is commonly used in regression models.[1] It is a value ranging from 0.00 (0%) to 1.00 (100%), with 0.00 meaning that the model explains none of the outcome, and 1.00 indicating a perfect fit.[2] It can be calculated using the variables mean of the observed Y values (Y'), Total Sum of Squares (TSS), and Residual Sum of Squares (RSS) [Figure 1].

If the R^2 value is closer to 1.0, then more of the fluctuation in the response (dependent) variable is strictly due to change in the predictor (independent) variable(s). Thus, the independent variable(s) in the model explain more of the variation of the dependent variable. Alternatively, as the R^2 value nears 0.0, it implies a weaker relationship between the independent variable(s) and the outcome.[3] This suggests that the model's predictions are not well-aligned with the actual data points. Thus, the independent variable(s) may not have much explanatory power concerning the variation observed in the dependent variable.[2] When the R^2 value is exactly 1.0, a perfect linear relationship exists between the independent variable(s) and the outcome. In this case, all data points fall exactly on a straight line, and the model provides a perfect prediction.[4] While this may seem ideal, it can also mean the model is "overfitting", which is a potential pitfall especially in the case of multivariate linear regression.[5] Put simply, overfitting can occur in two major ways. Firstly, the model is trained on a dataset that does not resemble the general population, and thus the model does not fit well when applied to new data. Second, if the model has more independent variables than data points, the model will necessarily always have a perfect fit, though this is unrealistic.[5] The latter point occurs

Calculating the R²value (coefficient of variance)

1. Calculate the mean of the observed Y values (Y')
   - Add all the observed Y values together and divide by the total number of observations.

2. Calculate the Total Sum of Squares (TSS)
   - TSS is the total amount of variability in the Y values concerning their mean Y'.
   - For each Y value, subtract the mean Y' (calculated in step 1), square the result, and sum up all the squared differences. TSS = Σ(Y - Y')²

3. Calculate the Residual Sum of Squares (RSS)
   - RSS is the variability that cannot be explained by the model
   - For each pair of Y and Ŷ values, subtract the predicted Ŷ value from the observed Y value, square the result, and sum up all the squared differences. RSS = Σ(Y - Ŷ)²

4. Calculate the R-squared value
   - It is the proportion of total variability in the Y values (TSS) that can be explained by the model (RSS).
   - R-squared = 1 - (RSS / TSS)

Figure 1. How to calculate R^2 value. Graphic designed by Latha Ganti on canva.com

because each variable can, in essence, be 'assigned' to a data point, such that the model will be able to perfectly "memorize" the dataset, resulting in a perfect fit for the training data.[6] However, this memorization does not generalize to new or unseen data. This is because the model essentially captures noise or random fluctuations in the training data, rather than underlying patterns or relationships. When tested on a new dataset, the model fails to accurately predict outcomes because the noise it captured is not present in the new data.[7] To avoid overfitting, techniques such as cross-validation, regularization, and reducing the number of variables can be employed. These approaches help ensure that the model captures general trends rather than specific quirks of the training data, improving its generalizability to unseen data.

In various fields, "good" R^2 values have different meanings. In the social sciences and psychology fields, where the behaviors of human subjects pose challenges, values as low as 0.10 to 0.30 are often considered acceptable.[8] The field of finance has a much larger range with "good" R^2 values ranging from 0.40 to 0.70,[9] depending on the nature of the analysis and data availability. Physical sciences and engineering generally expect higher R-squared values, above 0.70 to be considered good. Scientists in physics and chemistry generally consider 0.70–0.99 a "good" R^2 value.[10] Pure mathematics doesn't directly apply R^2 values, but when relevant, values should be close to a perfect 1.00 to indicate a data-model fit. In ecology, R^2 values can vary; values from 0.20 to 0.50 are considered acceptable or good, tailored to the specific research question and ecological context.[11] The field of medicine, however, does not have much conclusive data on this topic and researchers often use arbitrary bounds.[12]

REVIEW

INTERPRETING R^2 IN THE CONTEXT OF CLINICAL MEDICINE

To establish a benchmark for a 'good' R-squared value, a comprehensive review of the medical literature was conducted. In "Quantifying health", Dr. Chouiry examined over 43,000 papers in PubMed and noted that only a third of the papers that used linear regression even reported the R^2 value.[13] Therefore over 66% of published studies that utilize linear regression do not report this key statistic. The distribution of the R^2 values in the papers reviewed had a bimodal distribution, with 10% of papers having an R^2 < 0.035. One of the most significant findings was that the value of the R^2 (including high and low outliers) had no correlation to the impact factor of the journal. To place this information in the context of medicine, we looked at studies that informed a few of the bread-and-butter acute diagnoses seen in the emergency department: cardiac arrest, stroke, sepsis and head injury, with the idea that a narrower range of typical R^2 values may emerge.

CARDIAC ARREST

Out-of-hospital cardiac arrest is a condition with high mortality and poor outcomes even in settings where extensive emergency care resources are available.[14] A 2020 study of pediatric cardiac arrest in the United States found that predictors of survival to hospital discharge included female sex, number of minutes from collapse to the arrival of EMS, age (in months), and use of an advanced airway to be predictors in a regression model with an R^2 of 0.245.[15] Similarly, an adult cardiac arrest study from Croatia reported a logistic regression analysis for the return of spontaneous circulation (ROSC) to hospital admission, which included the following 5 factors: age, sex, adrenaline use, rhythm conversion and bystander CPR. This model had an R^2 of 0.217.[16]

INTRACEREBRAL HEMORRHAGE

Intracerebral hemorrhage (ICH) is the most devastating form of stroke. According to the World Stroke Organization, There are over 3.4 million new ICH cases each year. Globally, over 28% of all incidental strokes are intracerebral hemorrhages.[17,18] In a study comparing the ICH scores between men and women at ED arrival, the regression model comprised 16 factors including age, race sex, atrial fibrillation, hypertension, dyslipidemia, diabetes, smoking, independence in activities of daily living (ADLs), median BMI, median systolic and diastolic blood pressures, median hemoglobin A1C (HbA1C), median Glasgow Coma Score (GCS), median NIHSS, and arrival method. Each individual factor was not necessarily significant in univariate analysis. The overall multivariate model had a R2 of 17%.[19]

A study of patients undergoing cranioplasty (CP) after decompressive craniectomy following ICH examined the predictive factors for procedural complications and found that a history of primary coagulopathy, intraoperative ven-

tricular puncture, and intraoperative dural limit violation were associated with increased surgical complications. It additionally found that patients who lived at home at the time of CP had a reduced likelihood of post-CP surgical complications. This model explained 20% of the variance in post-CP complications (R^2=0.20).[20]

SEPSIS

Sepsis is the primary cause of death due to infection.[21] It is one of the most frequent causes of death worldwide, with 48.9 million cases and 11 million sepsis-related deaths worldwide, representing 20% of all global deaths.[22] A prospective study of heart rate variability as a predictor of mortality in sepsis reports on a model with an R^2 of 0.167. The Sequential Organ Failure Assessment (SOFA) Score were the elements in the model and included: FiO_2 level, mechanical ventilation, and platelet count.[23]

A 2020 study examined the appropriateness of empirical antibiotics in patients with sepsis in the ICU. The simple linear regression indicated that appropriate empirical antibiotics were associated with decreasing ICU length of stay, with a model R^2 of 0.055. By contrast, the use of inappropriate antibiotics was associated with a worse APACHE-II score, with an R^2 of 0.079.[24]

HEAD INJURY

Traumatic brain injury (TBI) is a prevalent condition with over 214,110 TBI-related hospitalizations in 2020 and 69,473 TBI-related deaths in 2021 in the United States alone.[25] The most common etiologies are falls and road traffic accidents.[26]

A retrospective observational study of over 800 patents examined the impact of seatbelt in motor vehicle associated head injury. Model variables included sex, alteration or loss of consciousness, vomiting, consumption of alcohol before the accident, and arrival via EMS. Three separate models with the outcomes of TBI severity, abnormal brain CT and ICU admission were built, with R2 values of 0.261, 0.228, and 0.208 respectively.[27]

The prospective longitudinal observational Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) study examined factors associated with good Glasgow Outcome Score 6 months post injury. The model included age, Glasgow Coma Score, and injury severity score (ISS), resulting in a model with an R^2 of 0.18. Adding sex, American Society of Anesthesiologists Physical Status (ASA Class), psychiatric history, cause of injury, and pupillary reactivity, to the model resulted in an R^2 of 0.21.[28]

SUMMARY OF FINDINGS

The above studies cover various pathologies, across various journals and disciplines, and converge around an R^2 of ~20%. Given that much of clinical medicine is impacted by humans' psychosocial underpinnings, a value of >15% is likely to be a reasonable threshold, providing that each of the individual variables in the model are significant[4] [Figure 2].

Understanding the limitations of the R^2 value in clinical medicine is critical. Medical outcomes are most often formed as a culmination of numerous complex factors, and relying solely on an R^2 value could oversimplify these nonlinear relationships. When dealing with limited sample sizes and large numbers of predictors, a risk of overfitting

What's a good R² value in Clinical Medicine?

>15%
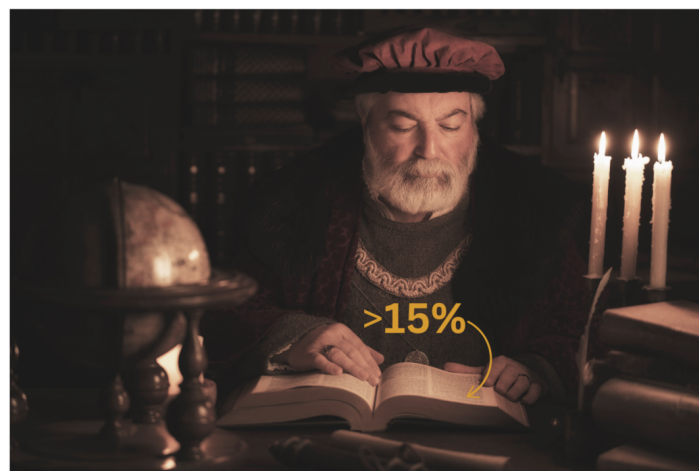
- Certain areas within clinical research, like studying complex disease interactions, may naturally have lower R² values compared to studies with more predictable outcomes.

- Even with a moderate R² value, if the effect size is clinically significant, the findings can still be valuable

- A low R² doesn't always mean a poor fit, and a high R² doesn't always mean a good fit.

Figure 2. What is a good R^2 value in clinical medicine? Graphic designed by Latha Ganti on canva.com.

exists, which inflates R-squared values and compromises the model's generalizability to new data.[6] Researchers should utilize R-squared alongside other statistical tests to avoid biases and inaccuracies.

"What is a good R-squared value in clinical medicine" is the classic question that begets the answer "it depends." Occasionally there is value in explaining only a very small fraction of the variance, particularly when outcomes are multifactorial such as the natural history of a disease.[29] The interpretation of R-squared values should be carefully considered within the specific context of the study design and the characteristics of the investigated population. While higher R-squared values may suggest a better fit of the regression model to the data, the intricacies of clinical medicine warrant a cautious approach to interpreting such values. The establishment of a 'good' R-squared benchmark should be further refined through a meticulous review of additional literature to account for the diversity and complexity of clinical research studies. Perhaps the best guideline in interpreting the R^2 for a particular research question is to base the target R^2 value off of similar studies in the literature. As the benchmarks vary widely in clinical medicine, comparison with existing literature is paramount in understanding the significance of an R^2 value that one may obtain in their own analysis.[30]

CONCLUSION

Through comparison of the data presented in the referenced articles, it is reasonable to consider an R-squared value of 0.15-0.20 (15-20%) as a suitable benchmark in clinical research, with many caveats. It is essential to acknowledge that the appropriateness of an R-squared benchmark can vary depending on the number and influence of factors contributing to the outcomes under investigation. Different studies may require specific R-squared thresholds to accurately gauge model performance and predictive capacity. Above all, R^2 benchmarks vary widely by the particular research question being addressed, and the most accurate benchmark will always be achieved through comparison with existing literature on the specific research question being investigated. It is important to note that no R^2 value tells you if a model is good or not; rather it tells you how well the model fits the data you have. To ensure robust and reliable results, researchers are encouraged to complement R-squared analysis with other relevant statistical tests and validation techniques. This comprehensive approach contributes to a more comprehensive evaluation of the regression model and enhances confidence in the study findings.

REFERENCES

1. Andrade C. Regression: Understanding What Covariates and Confounds Do in Adjusted Analyses. *J Clin Psychiatry*. 2024;85(4):24f15573. [doi:10.4088/JCP.24f15573](doi:10.4088/JCP.24f15573)

2. What is R-Squared? Accessed October 18, 2024. https://corporatefinanceinstitute.com/resources/data-science/r-squared/

3. Camirand Lemyre F, Chalifoux K, Desharnais B, Mireault P. Squaring Things Up with R2: What It Is and What It Can (and Cannot) Tell You. *J Anal Toxicol*. 2022;46(4):443-448. doi:10.1093/jat/bkab036

4. Hamilton DF, Ghert M, Simpson AH. Interpreting regression models in clinical outcome studies. *Bone Joint Res*. 2015;4(9):152-153. [doi:10.1302/2046-3758.49.2000571](doi:10.1302/2046-3758.49.2000571)

5. Montesinos López OA, Montesinos López A, Crossa J. Chapter 4, Overfitting, Model Tuning, and Evaluation of Prediction Performance. In: *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer; 2022. [doi:10.1007/978-3-030-89010-0_4](doi:10.1007/978-3-030-89010-0_4)

6. Gygi JP, Kleinstein SH, Guan L. Predictive overfitting in immunological applications: Pitfalls and solutions. *Hum Vaccin Immunother*. 2023;19(2):2251830. [doi:10.1080/21645515.2023.2251830](doi:10.1080/21645515.2023.2251830)

7. Charilaou P, Battat R. Machine learning models and over-fitting considerations. *World J Gastroenterol*. 2022;28(5):605-607. [doi:10.3748/wjg.v28.i5.605](doi:10.3748/wjg.v28.i5.605)

8. Chicco D, Warrens MJ, Jurman G. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021;7:e623. doi:10.7717/peerj-cs.623

9. Ozili PK. The Acceptable R-Square in Empirical Modelling for Social Science Research. *Social Research Methodology and Publishing Results*. Published online June 5, 2022. doi:10.2139/ssrn.4128165

10. R-Squared: Definition and How to Calculate. Accessed October 18, 2024. https://www.investopedia.com/terms/r/r-squared.asp

11. Lin Y, Wiegand K. Low R2 in ecology: Bitter, or B-side? *Ecological Indicators*. 2023:153. doi:10.1016/j.ecolind.2023.110406

12. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128-138. [doi:10.1097/EDE.0b013e3181c30fb2](doi:10.1097/EDE.0b013e3181c30fb2)

13. What is a Good R-Squared Value? [Based on Real-World Data]. Accessed October 18, 2024. https://quantifyinghealth.com/r-squared-study

14. Werner K, Hirner S, Offorjebe OA, et al. A systematic review of cost-effectiveness of treating out of hospital cardiac arrest and the implications for resource-limited health systems. *Int J Emerg Med*. 2024;17:151. [doi:10.1186/s12245-024-00727-w](doi:10.1186/s12245-024-00727-w)

15. Banerjee P, Ganti L, Stead TG, Vera AE, Vittone R, Pepe PE. Every one-minute delay in EMS on-scene resuscitation after out-of-hospital pediatric cardiac arrest lowers ROSC by 5. *Resusc Plus*. 2020;5:100062. doi:10.1016/j.resplu.2020.100062

16. Lovaković J, Šantek P, Mahečić LM, et al. Rhythm conversion in out-of-hospital cardiac arrest and influence on the return of spontaneous circulation at the hospital arrival: a 10-year retrospective study in Croatia. *Int J Emerg Med*. 2024;17:157. [doi:10.1186/s12245-024-00746-7](doi:10.1186/s12245-024-00746-7)

17. World Stroke Organization (WSO). Global Stroke Fact Sheet 2022. [doi:10.1177/17474930211065917](doi:10.1177/17474930211065917)

18. Ward S, Jamal BC, Ganti L. Frontoparietal intraparenchymal hemorrhage secondary to anticoagulation. *Int J Emerg Med*. 2024;17:130. [doi:10.1186/s12245-024-00723-0](doi:10.1186/s12245-024-00723-0)

19. Ganti L, Shameem M, Houck J, et al. Gender disparity in stoke: Women have higher ICH scores than men at initial ED presentation for intracerebral hemorrhage. *J Natl Med Assoc*. 2023;115(2):186-190. doi:10.1016/j.jnma.2023.01.013

20. Ferreira A, Viegas V, Cerejo A, Silva PA. Predictive factors for cranioplasty complications - A decade's experience. *Brain Spine*. 2024;4:102925. [doi:10.1016/j.bas.2024.102925](doi:10.1016/j.bas.2024.102925)

21. Banerjee A, Barbera A, Melton J, Banerjee P. Shock index for prehospital sepsis recognition. *Academic Medicine & Surgery*. Published online February 29, 2024. [doi:10.62186/001c.94556](doi:10.62186/001c.94556)

22. Sepsis. Accessed October 18, 2024. https://www.who.int/news-room/fact-sheets/detail/sepsis

23. de Castilho FM, Ribeiro ALP, da Silva JLP, Nobre V, de Sousa MR. Heart rate variability as predictor of mortality in sepsis: A prospective cohort study. *PLoS One*. 2017;12(6):e0180060. doi:10.1371/journal.pone.0180060

24. Al-Sunaidar KA, Prof. Abd Aziz N, Prof. Hassan Y. Appropriateness of empirical antibiotics: risk factors of adult patients with sepsis in the ICU. *Int J Clin Pharm*. 2020;42:527-538. [doi:10.1007/s11096-020-01005-4](doi:10.1007/s11096-020-01005-4)

25. Centers for Disease Control and Prevention. National Center for Health Statistics: Mortality Data on CDC WONDER. Accessed October 18, 2024. https://wonder.cdc.gov/mcd.html

26. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Care Services, Board on Health Sciences Policy, Committee on Accelerating Progress in Traumatic Brain Injury Research and Care. The Scope and Burden of Traumatic Brain Injury. In: Matney C, Bowman K, Berwick D, eds. *Traumatic Brain Injury: A Roadmap for Accelerating Progress*. National Academies Press (US); 2022. https://www.ncbi.nlm.nih.gov/books/NBK580076/

27. Ganti L, Bodhit AN, Daneshvar Y, et al. Effectiveness of seatbelts in mitigating traumatic brain injury severity. *World J Emerg Med*. 2021;12(1):68-72. doi:10.5847/wjem.j.1920-8642.2021.01.011

28. Mikolić A, Steyerberg EW, Polinder S, et al. Prognostic Models for Global Functional Outcome and Post-Concussion Symptoms Following Mild Traumatic Brain Injury: A Collaborative European NeuroTrauma Effectiveness Research in Traumatic Brain Injury (CENTER-TBI) Study. *J Neurotrauma*. 2023;40(15-16):1651-1670. [doi:10.1089/neu.2022.0320](doi:10.1089/neu.2022.0320)

29. Linear regression models. Accessed October 18, 2024. https://people.duke.edu/~rnau/rsquared.htm

30. Frasca M, La Torre D, Pravettoni G, et al. Explainable and interpretable artificial intelligence in medicine: a systematic bibliometric review. *Discov Artif Intell*. 2024;4:15. [doi:10.1007/s44163-024-00114-7](doi:10.1007/s44163-024-00114-7)