JBE Journal of Biostatistics and Epidemiology
J Biostat Epidemiol. 2025;11(1): 115-132

ORIGINAL ARTICLE
DOI: *https://doi.org/10.18502/jbe.v11i1.19321*
*Received **December** 16, 2024; Accepted **February** 13, 2025;*
*Published online **March** 15, 2025.*

Robust Neighborhood Confidence Interval and Width to Evaluate the Outcome of a Binary Random Variable of Unequal Cluster Sizes

Tareef **Fadil Raham**[1]*, Zaher **Fadhil Raham**[2], Abdulkhaleq Abduljabbar **Ali Ghalib Al-Naqeeb**[3]

[1]Department of Pediatrics, University of Warith Al-Anbiyaa, Baghdad, Iraq.
[2]Al-Nahrain Renewable Energy Research Center, Al-Nahrain university, Baghdad, Iraq.
[3]Medical & Health Technology College, Baghdad, Iraq.

ABSTRACT

Introduction: Confidence intervals (CIs) provide a more precise evaluation of outcomes, especially when the risk of an event is influenced by cluster size. While confidence intervals are commonly used to assess uncertainty in future data, in this study, we focus on their role in quantifying variability within currently observed outcomes. Specifically, the width of the predicted confidence interval serves as an indicator of existing intra-cluster heterogeneity, highlighting the extent of variability across different cluster sizes.
This study introduces a novel method for evaluating observed outcomes of dichotomous random variables in datasets with unequal binary cluster sizes. By employing a robust neighborhood confidence interval width, this approach ensures a more reliable and adaptive estimation of intra-cluster variability, allowing for a more accurate interpretation of current data distributions.

Methods: We introduce a novel algorithm for constructing an intra-cluster robust neighborhood confidence interval and its corresponding width for each cluster. This method enables the ranking of clusters based on confidence interval width, from the narrowest to the widest, providing a systematic approach to quantifying intra-cluster variability. By evaluating observed values within these ranked intervals, the algorithm offers a more precise assessment of data heterogeneity. To illustrate the effectiveness of this method, we present both a simulated example assessing its finite-sample performance and a real-world application in the context of antimicrobial resistance data with unequal binary cluster sizes.

Results: The robust neighborhood intra-cluster confidence interval (CI) width was successfully derived for interpreting binary outcome data with unequal cluster sizes. The analysis showed that narrow confidence intervals indicate minimal random variation, suggesting higher reliability in the observed results, whereas wider intervals highlight increased intra-cluster variability.

Conclusion: The intra-cluster robust neighborhood CI and its corresponding width provide a valuable tool for analyzing binary outcome data with unequal cluster sizes. This method enhances the interpretation of observed results by systematically quantifying variability within clusters, allowing for more reliable intra-cluster comparisons.

Key words: Dichotomous variable; Binary variable; Robust neighborhood confidence interval; Confidence interval width; Unequal cluster sizes

*Corresponding Author:
tareef.fadhil@uowa.edu.iq

INTRODUCTION

A dichotomous (binary) variable is a categorical data type that can take exactly two possible values. Unequal cluster (group) sizes with binary outcomes are common in cluster randomized trials (CRTs) because the number of participants across clusters varies in scientific research. Unequal cluster (group) sizes with binary outcomes are common type of outcome across clusters varies in scientific research.[1] This issue has been widely recognized in the literature, as CRTs are often used when individual randomization is impractical for logistical, ethical, or feasibility reasons. CRTs with small or medium cluster sizes pose a higher risk of type I errors due to inflated intra-cluster standard errors, which can lead to misinterpretations of treatment effects if not properly accounted for.[2] Studies have highlighted the importance of appropriate statistical methods to address unequal cluster sizes and ensure accurate inference. Standard approaches that fail to adjust for intra-cluster correlations may inflate Type I errors, leading to false-positive results. This raises concerns concerning interpretations, analytic approaches and methods that do not account for intra-cluster standard errors of effects that inflate type I errors.[3,4]

Addressing these challenges requires robust statistical methods that appropriately account for intra-cluster variability in binary outcomes. Researchers have explored various strategies to mitigate these risks, including adjusted standard error calculations, hierarchical models, and specialized statistical tests that ensure more reliable interpretations of CRT findings.

Equal sample sizes provide greater statistical power compared to unequal sample sizes. As the imbalance between cluster sizes increases, the statistical power decreases, making it more challenging to detect true effects. This occurs because unequal sample sizes lead to greater variance and reduced efficiency in estimating treatment effects, particularly in cluster randomized trials (CRTs) and other study designs involving hierarchical data structures.[5]

The impact of unequal cluster sizes on statistical power for binary outcomes has been extensively studied. Research indicates that power can be significantly affected when cluster sizes are highly variable, particularly in settings with large clusters and increased variability. Studies suggest that when intra-cluster variability exceeds 0.23, specialized design or analytical methods should be implemented to properly account for the imbalance in cluster sizes, ensuring accurate effect estimation and maintaining statistical power.[6-8]

However, when cluster size variability is below 0.3, adjustments are generally not necessary, as statistical power remains largely unaffected. In such cases, standard analytical methods can be applied without significant risk of inflated Type I errors or reduced power.[2,8,9]

Properly accounting for cluster size variation is essential to ensure accurate effect estimation and maintain the reliability of statistical inferences in study designs.[2]

The coefficient of variation (CV) of cluster size is determined by the mean cluster size and standard

deviation, as it represents the ratio of the standard deviation of cluster sizes to the mean cluster size. A higher CV indicates greater variability in cluster sizes, which can impact statistical power, efficiency, and estimation accuracy in studies with unequal cluster sizes. Properly accounting for CV in study design and analysis helps ensure robust statistical inference.

Multiple approaches for unequal cluster sizes in binary data that adjust the sample size formulae have been proposed. The leading models include the following: marginal, random-effects , conditional, response-conditional, and within-cluster resampling.[10] The traditional marginal methods are observation-based, while within-cluster resampling is a cluster-based method that equally weights each cluster.[10] Nevertheless, differences in cluster sizes led to variation between the statistical methods used in analyses in terms of power, coverage, effects estimation , and significance.[11]

Commonly used models based on asymptotic approximations for analyzing binary outcomes not accounting for various ranges of ICC, CV, number of clusters, and average cluster sizes are subjected to inaccurate conclusions.[2]

Multiple approaches adjusting the sample size formulae for unequal size clusters with binary outcome on design effects or relative efficiency considerations have been proposed such as the mean cluster size or the maximum cluster size in the simple design,[2] adjusted design effect[12] weighted rank tests that depend on the variance of cluster size.[13] Kennedy-Shaffer and Hughes[14] proposed analytic formulae for CRTs with binary outcomes through sample size calculation for the stratified vs un-stratified designs.

Taljaard et al. developed a sample size formulae by incorporating the missing values of the data,[15] while Hoffman et al.[10] suggested using within-cluster resampling for analyzing data.

Raham et al.[16] adjust unequal cluster numbers to a predicted number equals to the maximum cluster number among clusters. This adjustment maintains validity as far as the risk for the outcome of interest is related to the cluster size.

These methods involve trans-cluster analyses of unequal cluster sizes, furthermore, there is an existing overall paucity of methodological papers on binary outcomes in scientific research.[2,6] There is a notable scarcity of published methodologies specifically addressing binary cluster analyses and interpretations. In scientific research, methodological studies on binary outcomes remain limited. Most existing approaches focus on trans-cluster analyses involving unequal cluster sizes, whereas methods tailored for within-cluster (intra-cluster) analysis are largely under explored.

Variability reflexes unexplained variation, systematic error/s , and random effect/s.[17] The intra-cluster correlation coefficient (ICC) is influenced by multiple factors, including the number of clusters, cluster size, cluster size variability, event rate, and event rate variations. Ignoring the effects of unequal cluster sizes in data analysis can lead to estimation bias, inflated Type I error rates, and reduced statistical power and efficiency. In this context, we proposed a robust neighborhood

confidence interval (RNCI) estimation method for intra-cluster confidence intervals (CIs) in binary data with unequal cluster sizes. By this novel method, estimations are performed to evaluate observed outcomes of dichotomous random variables in unequal binary cluster sizes. This method focuses on ensuring that the variability in cluster sizes does not bias the confidence interval estimates, making comparisons across clusters more reliable.

Examining the width of the intra-cluster predicted interval as a function of the confidence level and variability will be a new method which provides a structured way to estimate robust confidence intervals (CIs) while accounting for cluster size variability in binary data with unequal cluster sizes.

METHODOLOGY

A proposed method aim is to handles unequal cluster sizes by adjusts for variability in cluster sizes by computing proportion-based estimates.

For predicting values of intra-cluster confidence intervals (CI) and widths in binary unequal sizes data and estimation a robust neighborhood confide interval is used to evaluate observed outcomes of dichotomous random variables. This approach helps to account for cluster size variability and ensures robust comparisons across clusters by estimating CIs that account for unequal cluster sizes. We proposed the following algorithmic steps:

Step 1: Calculate the rates of the 1st and 2nd outcomes of the observed binary frequencies in each cluster.

For each cluster, compute the rates of the first and second outcomes (i.e proportions) based on observed binary frequencies.

$$p = \frac{\text{number of positive cases}}{\text{total sample size}}$$

Step 2: Estimate significance boundaries: Sequentially estimate the minimum and maximum rates that achieve the desired level of significance (e.g., 0.05 for a 95% confidence level) using certain statistical method such as a two-sided Z-test or another test on the difference between proportions.

Z-test:

$$Plower = p - Z_{\propto/2} \times \sqrt{\frac{p(1-p)}{n}}$$

$$Pupper = p - Z_{\propto/2} \times \sqrt{\frac{p(1-p)}{n}}$$

Step 3: Compute count-based CI bounds: For the lower bound of the CI, multiply the minimum rate (from Step 2) by the total number of outcomes in the cluster.

For the upper bound of the CI, multiply the maximum rate (from Step 2) by the total number of outcomes in the cluster.

Lower Bound (LB)=p min ×n
Upper Bound (UB)=pmax ×n

Step 4: Construct percentage-based confidence intervals: The predicted confidence interval (CI) is formed by using the minimum estimate as the lower bound and the maximum estimates as the upper bound. This interval captures the predicted variability in proportions with the desired confidence level(e.g., 95%).

Step 5: Compute the width of the CI:
Percentage-based CI Width=UB−LB
Evaluate outcome values: Compare the predicted outcome values (from Step 4) based on the differences between the maximum and minimum rate values. The lower width (narrower CI) is considered the best or most reliable estimate.

Step 6: Interventions are re-ranged in two-direction ranks in descending orders according to estimated distances independently.
Rank the clusters independently in descending order based on the estimated distances between the confidence interval bounds.
This allows for comparison across different clusters, ensuring that interventions with more reliable estimates (i.e., narrower CIs) receive priority.

Step 7: Compare the observed CI widths across clusters. Identify clusters with consistently narrow intervals, indicating greater precision. Interventions in clusters with wider CIs may require further investigation or adjustments.

APPLICATION AND RESULT

We took an example of 15 different protocols applied for a certain medical condition. The number of participants differs among these groups (clusters) and the data fulfill the criteria of dichotomous (binary) nominal data of unequal sizes.

The observed frequencies: Positive (cured), and negative (not cured) are shown in Table 1. For predicting values of CIs at a p-value of 0.05 we use the propose steps already mentioned in the methodology section.

Using a two-sided Z-test on the difference between proportions, determine the minimum and maximum rate estimates that satisfy a given significance level (e.g., 0.05 for a 95% confidence interval).

The Z-score for a 95% CI is typically 1.96. Ensure that the estimated boundaries remain within valid probability ranges (0 to 1). According to this pvalue the upper and lower borders of predicted CI were exactly the 95% CI.

Table 1. Distribution of constrain simulated data according to observed frequencies: positive (cured), and negative (not cured) outcomes with their predicted CI s at P-value = 0.05 according to the difference between two proportions test of (two-sided Z-test)

Protocol	Positives (cured)				Negatives (not cured)				Total number
	Observed		Predicted (p-value at 0.05)		Observed		Predicted (p-value at 0.05)		
	No.	%	L.b:	Ub:	No.	%	L.b:	Ub:	
Protocol 1	21	25.93	0.1365 (11.057)	0.4040 (32.724)	60	74.07	0.5960 (48.276)	0.8635 (69.944)	81
Protocol 2	51	100	0.92724 (47.28924)	1.00 (51)	0	0.00	0(0.00)	0.07276 (3.71076)	51
Protocol 3	69	100	0.9461 (65.2809)	1.00 (69)	0	0.00	0(0.00)	0.0539 (3.7191)	69
Protocol 4	6	8.33	0.0132 (0.9504)	0.1965 (14.1480)	66	91.67	0.8831 (63.5832)	0.9868 (71.0496)	72
Protocol 5	3	9.09	0.04859 (1.60347)	0.14435 (4.76355)	30	90.91	0.85585 (28.24305)	0.9510 (31.3830)	33
Protocol 6	75	94.94	0.85695 (67.69905)	0.99885 (78.90915)	4	5.063	0.00100 (0.079)	0.1430 (11.297)	79
Protocol 7	2	0.00	0 (0.00)	0.03936 (3.77856)	96	100	0.96065 (92.2224)	1.00 (96)	96
Protocol 8	6	8.33	0.0132 (0.9504)	0.1965 (14.148)	66	91.67	0.8831 (63.5832)	0.9868 (71.0496)	72
Protocol 9	6	33.33	0.070685 (1.27233)	0.66064 (11.89152)	12	66.7	0.34014 (6.12252)	0.92925 (16.7265)	18
Protocol 10	9	75.0	0.3525 (4.2300)	0.97885 (11.7462)	3	25.0	0.02037487 (0.244498)	0.580943 (6.971316)	12
Protocol 11	36	92.3	0.7615 (29.6985)	1.00 (36.0)	3	7.69	0(0.00)	0.2385 (9.3015)	39
Protocol 12	21	87.5	0.6316 (15.1584)	1.000 (24.0)	3	12.5	0(0.00)	0.3686 (8.8464)	24
Protocol 13	9	100	0.64868 (5.83812)	1.00 (9)	0	0.00	0(0.00)	0.35132 (3.16188)	9
Protocol 14	14	100	0.7234 (8.6808)	1.00 (12)	0	0.00	0(0.00)	0.2766 (3.3192)	12
Protocol 15	0	0.00	0 (0.00)	0.35132 (3.16188)	9	1.00	0.64868 (5.83812)	1.00 (9)	9

The result is not signigicant at P=0.05. Z id 1.96 (two sided).

Table 1 shows the distribution of constraint simulated data according to observed frequencies: positive (cured), and negative (not cured) outcomes with their predicted CI s at Pvalue = 0.05 according to the difference between two proportions test of (two sided Z-test).

Table 2. Re-evaluating positive (cured) outcomes according to predicted CI widths of studied protocols, corresponding observed positive results % and sample size orders

Protocols	Sample size from high to low	Sample size order rank from high to low	Predicted positives (cured) predicted (p-value at 0.05)		% observed positives (cured)	Observed positives order rank
			Difference	Precision order rank from narrow to wide		
Protocol 7	96	1	0.039	1	0.00	14.5
Protocol 3	69	6	0.054	2	100	2.5
Protocol 2	51	7	0.073	3	100	2.5
Protocol 5	33	9	0.096	4	9.09	11
Protocol 6	79	3	0.142	5	94.94	5
Protocol 4	72	4.5	0.183	6.5	8.33	12.5
Protocol 8	72	4.5	0.183	6.5	8.33	6
Protocol 11	39	8	0.239	8	92.3	10
Protocol 1	81	2	0.268	9	25.93	2.5
Protocol 14	12	13	0.277	10	100	2.5
Protocol 13	9	14.5	0.351	11.5	100	14.5
Protocol 15	9	14.5	0.351	11.5	0.00	14.5
Protocol 12	24	10	0.368	13	87.5	7
Protocol 9	15	11	0.590	14	33.33	9
Protocol 10	12	12	0.626	15	75.0	8

Table 2 shows precision of outcome and rank according to predicted CI widths and corresponding observed positive results % and sample size orders. According to the constraint simulation example precision order rank (from narrow to wide) differs from sample order rank size in most of protocols. The precision order rank with a correspondence sample size order is 1 for 1st, 2 for 6th . 3 for 7th, 4 for 9th, 5 for 3^{d}, 6.5 for 4.5th, 8 for 8th, 9 for 2nd, 10 for 13th, 11.5 for 14.5th, 13 for 10th ,14 for 11th , and 15 for 12th. The protocol 10 with a relatively high observed positive results (cure) results of (75%) with a lowest precision order in our example despite the sample size order is not the lowest. The protocol 13 is better than protocol 10 in cure rate (100% vs 75%) despite sample size is larger in protocol 10 (12 vs 9). This indicate the significance of CI width compared to sample size in predicting the precision of outcome.(table 2)

Table 3 shows the same findings for CI difference and precision for negativity results for each corresponding protocol.

Table 4 shows application of this novel algorythm on a real data on antimicrobial resistance derived from previous study.16. Ceftriaxone, Piperacillin, and Cefepime show moderate variations

Table 3. Re-evaluating negative (not cured) outcomes according to predicted CI widths of studied protocols, corresponding observed negative results % and orders, and sample size orders.

Protocols	Sample size from high to low	Sample size order rank from high to low	Predicted negatives (not cured) predicted (p-value at 0.05)		Observed negatives (not cured)%
			CI difference	Precision order from narrow to wide	
Protocol 7	96	1	0.039	1	100
Protocol 3	69	6	0.054	2	0
Protocol 2	51	7	0.073	3	0
Protocol 5	33	9	0.096	4	90.91
Protocol 6	79	3	0.142	5	5.063
Protocol 4	72	4.5	0.183	6.5	91.67
Protocol 8	72	4.5	0.183	6.5	91.67
Protocol 11	39	8	0.239	8	7.69
Protocol 1	81	2	0.268	9	74.07
Protocol 14	12	13	0.277	10	0
Protocol 13	9	14.5	0.351	11.5	0
Protocol 15	9	14.5	0.351	11.5	1
Protocol 12	24	10	0.368	13	12.5
Protocol 9	18	11	0.590	14	66.7
Protocol 10	12	12	0.626	15	25

Table 3 Shows the same findings for CI difference and precision for negativity results for each corresponding protocol.

Table 4. Shows the real data related CI difference and precision for resistance of isolates results for each corresponding antibiotic

Protocol 7	Observed Resistance (No.)	Observed Resistance (%)	Total Isolates	Observe d Resistance Proportion	CI Lower Bound	CI Upper Bound	CI Width	Precision Rank	Resistance Rank
Ceftazidime	23	100	23	1	1	1	0	1	1
Cefotaxime	17	100	17	1	1	1	0	1	1
Tobramycin	3	100	3	1	1	1	0	1	1
Amikacin	4	100	4	1	1	1	0	1	1
Ceftriaxone	25	96.15	26.00 10400 41601 662	0.9615	0.887 54531 01023 066	1	0.112 45468 98976 9336	5	5
TMP-SMX	2	8.33	24.00 96038 41536 613	0.0833	0	0.193 83496 81925 6069	0.193 83496 81925 6069	6	12
Azithromycin	2	8.33	24.00 96038 41536 613	0.0833	0	0.193 83496 81925 6069	0.193 83496 81925 6069	6	12
Cefepime	12	92.3	13.00 10834 23618 635	0.922999 9999999 999	0.778 08527 66932 68	1	0.221 91472 33067 32	8	6

Table 4 continues.

Protocol 7	Observed Resistance (No.)	Observed Resistance (%)	Total Isolates	Observe d Resistance Proportion	CI Lower Bound	CI Upper Bound	CI Width	Precision Rank	Resistance Rank
Meropenem	1	9.09	11.00 11001 10011 002	0.0909	0	0.260 77355 81117 7205	0.260 77355 81117 7205	9	11
Tetracycline	7	25.92	27.00 61728 39506 17	0.259200 0000000 0004	0.093 93056 76659 3047	0.424 46943 23340 6964	0.330 53886 46681 392	10	10
Gentamycin	7	87.5	8	0.875	0.645 82348 50600 961	1	0.354 17651 49399 039	11	7
Piperacillin	3	75	4	0.75	0.325 64755 21456 251	1	0.674 35244 78543 75	12	8
Tazocine (Piperacillin/ Tazobactam)	2	33.33	6.000 60006 00060 01	0.3333	0	0.710 47388 46464 989	0.710 47388 46464 989	13	9
Minocycline	0	0	0	0	0	1	1	14	14
Ciprofloxacin	0	0	0	0	0	1	1	14	14

between rankings.This suggests partial agreement between the two methods but some discrepancies in how each method defines expected resistance levels. It shows minimal or no rank differences with gentamycin and meropenem which had identical rankings in both methods, showing that both approaches agree on these antibiotics' relative resistance levels nevertheless, major rank differences (≥6 ranks apart) have been notice with amikacin (7 ranks apart) & tobramycin (6 ranks apart). The differences illustrates clearly how variability through confidence intervals can change the conclusions. Accounting for variability provides a statistical measure of precision via CI width. furthermore, more reliable ranking when achieved it minimizes uncertainty in small and different sample sizes.

Validation

Pearson correlation analysis was conducted to evaluate the statistical agreement between the real data observed and predicted resistance values.

Correlation coefficient was 0.920 (very strong positive correlation) and P-value: 0.000001 (highly significant) indicating that there is a strong statistical correlation between observed and predicted resistance values. This suggests that the introduced method successfully predicts resistance trends, as higher observed resistance rates correspond strongly to higher predicted resistance values.

Therefore, the novel method provides a statistically reliable approach to estimating and ranking antibiotic resistance trends. We have conducted a comparative analysis between our proposed method and traditional confidence interval estimation techniques, including: Standard wald Confidence Intervals – commonly used in clinical and epidemiological studies but does not adjust for unequal cluster sizes, potentially leading to biased estimates: Wilson Score Intervals – more precise in small samples but still does not fully address intra-cluster variability in cluster-randomized trials (CRTs): Bootstrap-Based Intervals – useful for handling complex data structures, yet computationally intensive and less efficient for practical application in large-scale studies: and Bayesian Credible Intervals – incorporates prior distributions for uncertainty estimation but requires subjective prior assumptions, which may not always be feasible in real-world clinical studies.

The RNCI method outperformed traditional methods in accounting for intra-cluster variability by adapting CI width to cluster size differences, ensuring more precise ranking of interventions.

DISCUSSION

The proposed method was developed to address the challenges posed by unequal cluster sizes in binary outcome data. By incorporating a statistical test and confidence interval estimation, this approach ensures robustness by accounting for sampling variability within each cluster.

The width of the confidence interval is a function of the following elements: Confidence level, sampling error, sample size, and variability. The confidence interval measure degrees of precision characterizing estimated points of interest. Narrow confidence intervals bounds suggest the results are not subjected to a high degree of random variations.

The number of participants enrolled across clusters in scientific research usually varies. Increasing the sample size is a primary way to reduce the widths of confidence intervals but this can’t be achieved most of the time for different reasons.

If the width of the predicted confidence interval (CI) is narrow, the observed outcome estimate is more likely to be valid, particularly in the absence of systematic error. Conversely, a wider interval indicates a larger margin of error, meaning the true parameter value is likely to fall within a broader range. Such extensive ranges reflect imprecise estimates and should be interpreted with caution when drawing conclusions.

The use of confidence intervals (CIs) is a powerful tool for interpreting results across various study designs. In clinical trials, where patients are treated different hospitals, robust CIs help prevent false conclusions about treatment effects. In public health research, when analyzing disease prevalence across regions, this approach ensures fair comparisons between geographical areas. Similarly, in economics and social sciences, where survey respondents may come from unequal group sizes, the use of robust confidence intervals helps provide unbiased policy recommendations by accounting for sample variability.

The confidence intervals will be broad when the sample standard deviation is high. The greater the confidence level, the wider the confidence level. The most commonly desired alpha value is p = 0.05, 0.1, 0.01, and even 0.00118 can be used with the proposed intra-cluster predicted replace with: RNCI width.

"The calculation of trans-cluster confidence intervals around the difference between two group means is a fundamental aspect of Schuirmann's equivalence test.[19] This statistical approach provides evidence for equivalence by demonstrating that two study groups can be considered statistically similar when their confidence intervals fall within a predefined equivalence margin. In the context of intra-cluster analysis, this principle supports the robustness of comparisons across clusters by ensuring that observed differences remain within acceptable statistical limits."[20,21]

The precision of outcome is related to reproducibility and repeatability, ie the degree to which repeated experiments or tests under unchanged conditions show the same result findings i.e. the predicted confidence interval.[22, 23]

As illustrated by constraint simulation example, orders of CI differences differ greatly from orders of sample sizes. The predicted robust neighborhood CI differences among protocols can't be explained by the differences in sample sizes alone. This signifies that protocols with narrow differences reveal have lesser outcome variability. Small sized clusters can have narrow CI width and vise versa. Furthermore , equal sized clusters can be evaluated according to the CI width. By this new method, the predicted robust neighborhood CI width will be informative in interpretation the observed outcome values. The predicted robust neighborhood cluster outcome is highly valid if has no systematic error (high accuracy) and has lesser variability (narrow predicted confidence interval).

Our RNCI method compare how much outliers or distribution of outcomes affect conclusions.The application of RNCI will help to get precise conclusions and recommendations concerning binary data of unequal cluster sizes. Results showed that small sized sample can possibly have narrow CI width and vice versa. For this reason this application has a place in analyzing data with unequal clusters in various fields including health-related fields, such as medicine, pharmacology, biology, public health and etc.

Compared to other neighborhood-based methods, our approach estimates a more robust confidence interval (CI) that precisely achieves the desired significance level by applying the difference between two proportions test. Unlike standard methods, which may not fully account for intra-cluster variability, our approach ensures statistical robustness by explicitly incorporating cluster size differences.

Other statistical techniques, such as the clustered bootstrap method, resample entire clusters rather than individual data points, preserving the intra-cluster structure. Sandwich estimators provide robust standard errors that adjust for heteroscedasticity and intra-cluster correlations, improving variance estimation. Additionally, Huber and Tukey loss functions offer a way to mitigate the impact of within-

cluster outliers by down-weighting large residuals during model estimation. While these methods enhance robustness in different contexts, our proposed approach directly focuses on accurately estimating intra-cluster confidence intervals with minimal bias.

Upon reviewing the literature, we have identified the following studies that, while not identical, share some similarities with this current research:

1- Comparison of Confidence Interval Methods for the Intraclass Correlation Coefficient in Cluster Randomized Trials with Binary Outcomes[24]: This study evaluates various confidence interval methods applied to binary outcomes in community intervention trials, focusing on the intraclass correlation coefficient (ICC).

2- New Improved Estimators for Overdispersion in Models with Clustered Multinomial Data and Unequal Cluster Sizes[25]: This research proposes new estimators addressing overdispersion in clustered multinomial data, specifically considering unequal cluster sizes.

3- ICCbin: Facilitates Clustered Binary Data Generation and Estimation of Intracluster Correlation Coefficient (ICC) for Binary Data[26]: This R package assists in generating clustered binary data and provides estimates of the ICC using various methods, including confidence interval estimation.

These studies offer methodologies and discussions that are relevant the analysis of clustered binary data with unequal cluster sizes but not the unique contribution and advantage of our method.

Strengths of the Algorith

The novel algorithm presents several advantages in statistical analysis, particularly in its ability to handle binary outcomes within unequal cluster sizes. A key strength is its ability to enhance comparability, allowing meaningful comparisons across clusters by ranking interventions based on confidence interval (CI) widths. This ranking system provides a clearer interpretation of intra-cluster variability, ensuring that interventions with more precise estimates are prioritized over those with greater uncertainty. Additionally, the algorithm strengthens decision-making by identifying clusters with the most reliable estimates, facilitating a data-driven approach to intervention prioritization. This ensures that conclusions are based on statistically sound assessments, reducing the risk of misinterpretation due to cluster size variability.

A major advantage of this approach is its strong statistical foundation. The algorithm utilizes standard statistical methods to compare proportions across different groups while incorporating confidence intervals to provide a range of likely outcomes rather than a single estimate. By ensuring that treatment effectiveness is interpreted within a confidence range, the method enhances reliability and precision. The systematic ranking of interventions based on CI width prioritizes more stable and consistent estimates, minimizing errors caused by variations in observed data. Notably, step 6 allows for a structured comparison of interventions, integrating both effectiveness and certainty—an essential feature in clinical decision-making.

Unlike many traditional statistical approaches that stop at confidence intervals or hypothesis testing,

this method extends beyond conventional analysis by refining confidence intervals and using them as a ranking criterion for systematic intervention comparison. While most ranking methods in clinical research rely on effect sizes such as odds ratios, hazard ratios, or mean differences, this algorithm introduces a novel approach by prioritizing interventions based on CI width. This innovation is particularly useful in studies involving multiple interventions, where ensuring the most statistically reliable estimates is essential.

Another key strength of the algorithm is its flexibility in statistical methods, allowing researchers to apply various analytical approaches depending on study design and data characteristics. It accommodates techniques such as the Z-test, which is suitable for large sample sizes, and the T-test, which is ideal for small sample sizes with unknown variance. The Chi-square test can also be used for categorical data analysis, while the bootstrap method provides a non-parametric alternative for estimating confidence intervals without strict distributional assumptions. Additionally, Bayesian methods can be employed to update prior knowledge, improving inference in limited sample sizes or highly variable datasets. More advanced techniques, such as generalized estimating equations (GEE) or mixed-effects models, enhance estimation accuracy by accounting for intra-cluster correlations.

The algorithm also addresses the challenge of wide confidence intervals, which can impact result interpretation and decision-making. Adjustments can be made to improve precision, such as increasing sample sizes, applying hierarchical models, or using variance-stabilizing techniques to narrow CI widths. By incorporating these adjustments, the algorithm ensures that confidence intervals more accurately reflect the underlying data structure, strengthening the reliability of statistical estimates and improving decision-making in epidemiology, clinical research, and public health studies.

This novel algorithm represents a significant advancement in statistical methodologies, integrating a robust ranking mechanism with flexible statistical modeling options. Its ability to account for intra-cluster variability, refine confidence intervals, and prioritize interventions based on CI width makes it a powerful tool for analyzing binary outcome data with unequal cluster sizes. The flexibility in statistical methodology enhances the robustness of the algorithm, making it applicable across various data distributions, cluster sizes, and outcome types. By allowing researchers to select the most appropriate statistical method, the algorithm ensures accurate intra-cluster confidence interval estimation, facilitating reliable comparisons and data-driven decision-making in clinical, epidemiological, and public health research.

This adaptability further enhances the precision of results, ensuring that confidence intervals more accurately represent the true variability in the data and improving overall decision-making reliability.

Limitations of the Algorithm

While the predicted RNCI method enhances the interpretation of findings in studies with binary outcomes and unequal cluster sizes, it has certain limitations. One of the key challenges is that the predicted CI contains the most compatible values, but it does not directly describe the probability of

including the true parameter value.[17] Confidence intervals are designed to provide a plausible values, but each interval either includes or excludes the true value, with the probability of any given interval being either 0 or 1.27 This inherent characteristic means that interpretation of uncertainty should be done with caution, particularly in studies with small sample sizes.

Another limitation lies in the fact that CI width alone may not fully capture reliability. While a narrower CI width generally indicates higher precision, it does not always guarantee greater accuracy. Small sample sizes, for instance, may lead to artificially narrow CIs, which increases the risk of overconfidence in the estimates. Additionally, the method does not explicitly adjust for sample size variability, meaning that protocols with smaller sample sizes may appear more precise but still suffer from statistical uncertainty. Future improvements could incorporate Bayesian approaches or adjusted confidence intervals, such as the Wilson score interval, to enhance reliability in small sample settings.

The algorithm also assumes that observed differences between interventions are purely due to the protocols themselves, without accounting for potential confounding factors. In real-world applications, various factors—such as differences in patient characteristics, healthcare settings, or external influences—could contribute to observed variations. The current method does not integrate covariate adjustments or hierarchical modeling, which could improve the accuracy of comparisons by controlling for study-specific biases.

While prioritizing interventions based on confidence interval width is an innovative approach, integrating multiple ranking criteria—such as combining CI width with effect sizes—could enhance the method's robustness in clinical decision-making.

Lastly, while the algorithm effectively handles binary outcome data, its application to multi-category or continuous outcomes remains unexplored. Despite these limitations, the RNCI method presents a novel and statistically rigorous approach for handling unequal cluster sizes in binary outcome studies. However, additional research is needed to validate its application in different study settings, particularly for small sample sizes, confounding factor adjustments, and multi-variable outcome modeling.

Future Consideration

Further validation studies are necessary to evaluate the consistency and reliability of this approach across diverse datasets and real-world applications. Future developments may incorporate Bayesian methods, hybrid ranking strategies, or machine-learning-driven refinements to enhance the precision and interpretability of confidence intervals, particularly in clinical and epidemiological research. Additionally, expanding the method to accommodate continuous data distributions or integrating it with hierarchical models, such as generalized estimating equations (GEE) or mixed-effects models, could improve its adaptability in more complex study designs and multivariable analyses.

CONCLUSION

The results from the constrained simulation example underscore the importance of CI width over sample size alone in evaluating the precision of observed outcomes. The re-evaluated ranking orders indicate that certain protocols with a high observed positive outcome received a lower precision ranking than expected, despite having a larger sample size. Conversely, a protocol with a smaller sample size ranked higher in precision. This demonstrates that CI width serves as a more reliable measure of precision compared to sample size alone, reinforcing the significance of this approach in ranking and evaluating interventions.

Furthermore, the real-world antimicrobial resistance example data analysis highlights the applicability of this method in evaluating resistance trends among bacterial isolates. The comparison between observed and predicted resistance rankings revealed significant discrepancies in certain antibiotics, where the ranking based on observed resistance did not align with the predicted CI with based ranking. This indicates that CI width could provide a more precise measure of reliability than direct resistance proportions emphasizing how CI width can refine interpretations of resistance data by accounting for statistical uncertainty.

These findings confirm that CI width is a crucial factor in determining the reliability of observed outcomes, whether in clinical interventions or bacterial resistance assessments. This reinforces the need for a ranking system that prioritizes precision over raw proportions, ensuring that statistical analyses yield more reliable conclusions. By applying this methodology to antimicrobial resistance studies for example , researchers can gain a better understanding of resistance patterns and improve the accuracy of intervention assessments in infectious disease research.

Recommendations

The intra-cluster predicted robust neighborhood confidence interval (CI) and its corresponding width represent a novel analytical tool for handling binary outcome data with unequal cluster sizes. This method provides a statistically rigorous approach to assessing intra-cluster variability, ensuring more reliable and precise comparisons in cluster-based analyses. Future research should focus on expanding its applicability to various study designs, refining its statistical framework, and incorporating adjustments for small-sample effects to further enhance its accuracy and interpretability in real-world research settings.

Author’s contributions

This work was carried out in the collaboration among TFR:

Find the idea of the research, contributed for the titled formulation, wrote introduction, carried out discussion of results, gathered the initial data, wrote introduction, carried out discussion of results, data management.

(Zaher Fadhil Raham): Find the idea of the research, carried out discussion of results. A A. A. G.A: Find the idea of the research, designed the study, and formulation the title of the research with the aim of the study, find the methodology and the algorithm of the research, analysis, and findings results for the application, data management.

All authors read and approved the final manuscript.

Conflicts

Authors declare that there is no conflicts of interests.

Funding

No funding personals or bodies had any role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Kahan BC, Forbes G, Ali Y, et al. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials. 2016;17:438. https://doi.org/10.1186/s13063-016-1571-2.

2. Zhan D, Xu L, Ouyang Y, Sawatzky R, Wong H. Methods for dealing with unequal cluster sizes in cluster randomized trials: A scoping review. PLoS ONE. 2021; 16(7): e0255389. https://doi.org/10.1371/journal.pone.0255389.

3. Thompson JA, Leyrat C, Fielding KL, Hayes RJ. Cluster randomised trials with a binary outcome and a small number of clusters: comparison of individual and cluster level analysis method. BMC Med Res Methodol. 2022 ; 12: 22(1):222. doi: 10.1186/s12874-022-01699-2. PMID: 35962318; PMCID: PMC9375419.

4. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. Am J Public Health. 2004;94(3):423-32. Doi: 10.2105/ajph.94.3.423. PMID: 14998806; PMCID: PMC1448268.

5. Rusticus, Shayna A. & Lovato, Chris Y. Impact of Sample Size and Variability on the Power and TypeI Error Rates of Equivalence Tests: A Simulation Study. Practical Assessment, Research & Evaluation. 2014; 19 (11). Available online: http://pareonline.net/getvn.asp?v=19&n=11.

6. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: Effect of coefficient of variation of cluster size and analysis method. Int. J. Epidemiol. 2006;35(5):1292–1300. pmid:16943232.

7. Braun TM. A mixed model formulation for designing cluster randomized trials with binary outcomes. Statistical Modelling. 2003;3(3):233–249.

8. Candel MJ, Van Breukelen GJ. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order pql mixed logistic regression. Stat. Med. 2010;29(14):1488–1501. pmid:20101669.

9. Jung SH, Kang SH, Ahn C. Sample size calculations for clustered binary data. Stat. Med. 2001;20(13):1971–1982. pmid:11427953.

10. Elaine B. Hoffman, Pranab K. Sen, Clarice R. Weinberg, Within-cluster resampling, Biometrika,. 2001; 88(4): 1121–1134, https://doi.org/10.1093/biomet/88.4.1121.

11. Durán Pacheco G, Hattendorf J, Colford JM Jr, Mäusezahl D, Smith T. Performance of analytical methods for overdispersed counts in cluster randomized trials: sample size, degree of clustering and imbalance. Stat Med.. 2009;28(24):2989-3011. Doi: 10.1002/sim.3681. PMID: 19672840.

12. Kong S-H, Ahn CW, Jung S-H. Sample size calculation for dichotomous outcomes in cluster randomization trials with varying cluster size. Drug Inf. J. 2003;37(1):109–114.

13. Li J, Jung SH. Sample size calculation for cluster randomization trials with a time-to-event endpoint. Stat. Med. 2020;39(25):3608–3623. pmid:33463748.

14. Kennedy-Shaffer L, Hughes MD. Sample size estimation for stratified individual and cluster randomized trials with binary outcomes. Stat. Med. 2020;39(10):1489–1513. pmid:32003492.

15. Taljaard M, Donner A, Klar N. Accounting for expected attrition in the planning of community intervention trials. Stat. Med. 2007;26(13):2615–2628. pmid:17068842.

16. Raham, TF, Ali Al. Zubaidi HH, Olewi AO, Abddul-Fatah Al-Aboosi AA et al. A Novel Algorithm for Predicting Antimicrobial Resistance in Unequal Groups of Bacteria Isolates. International Journal of Statistics in Medical Research. 2023 ;12,226-232. https://doi.org/10.6000/1929-6029.2023.12.26

17. Kümmel A, Bonate PL, Dingemanse J, Krause A. Confidence and Prediction Intervals for Pharmacometric Models. CPT Pharmacometrics Syst Pharmacol. 2018;7(6):360-373. Doi: 10.1002/psp4.12286. Epub 2018 Mar 25. PMID: 29388347; PMCID: PMC6027739.

18. Bevans R. (2023, June 22). Understanding Confidence Intervals | Easy Examples & Formulas. Scribbr. Retrieved December 8, 2023, from https://www.scribbr.com/statistics/confidence-interval/.

19. Schuirman D J. A comparison of the two one-sided tests procedures and the power

approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics. 1987; 15: 657-680.

20. Blackwelder WC. Current issues in clinical equivalence trials. Journal of Dental Research. 2004; 83:pp113- 115.

21. Rogers JL, Howard, KI & Vessey, JT. Using significance tests to evaluate equivalence between two experimental groups. Pyschological Bulletin. 1993;113: 553-565.

22. Taylor, John Robert (1999). An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books. pp. 128–129. ISBN 0- 935702-75-X.

23. Hartnack, S., Roos, M. Teaching: confidence, prediction and tolerance intervals in scientific practice: a tutorial on binary variables. Emerg Themes Epidemiol ; 2021:18, 17. https://doi.org/10.1186/s12982-021-00108-1.

24. Braschel MC, Svec I, Darlington GA, Donner A. A comparison of confidence interval methods for the intraclass correlation coefficient in community-based cluster randomization trials with a binary outcome. Clin Trials. 2016 Apr;13(2):180-7.

25. Alonso-Revenga JM. Martín N & Pardo L. New improved estimators for overdispersion in models with clustered multinomial data and unequal cluster sizes. Stat Comput 27, 193–217 (2017). https://doi.org/10.1007/s11222-015-9616-z.

26. Hossain A, ChakrabortyH. Facilitates Clustered Binary Data Generation, a Estimation of Intracluster Correlation Coefficient (ICC) for Binary Data.Version 1.1.1 . 2017.L https://cran.r-project.org/package=ICCbin.

27. Ashley I Naimi, Brian W Whitcomb. Can Confidence Intervals Be Interpreted?, American Journal of Epidemiology. 2020; 189 (7):631–633. https://doi.org/10.1093/aje/kwaa004.