mathematics

MDPI

Article

An Alternative to the Bland—Altman Repeated-Measures Correlation to Account for Variability of Slopes Across Persons

Tyler M. Moore [1,2,*] and Mathias Basner [3]

[1] Psychosis and Neurodevelopment Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, 3700 Hamilton Walk, Philadelphia, PA 19104, USA
[2] Lifespan Brain Institute (LiBI), Children's Hospital of Philadelphia and Penn Medicine, Philadelphia, PA 19104, USA
[3] Behavioral Regulation and Health Section, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
* Correspondence: tymoore@pennmedicine.upenn.edu

Abstract: The Bland–Altman repeated-measures correlation (rmcorr) is widely used to estimate within-person correlations between two variables in repeated-measures data. However, it assumes the same slope for all subjects, which can be misleading when slopes vary. We propose an alternative method, the weighted mean within-person correlation (wmcorr), which calculates the average of all within-person correlations, weighted by the square root of the number of observations for each person. The wmcorr method was applied to real data examples and compared to the Bland–Altman rmcorr. Simulations (5000) were run to compare the mean significance levels (p-values) of rmcorr and wmcorr and to determine the relationship between estimated rmcorrs and wmcorrs. In most cases, rmcorr and wmcorr yielded similar results. However, in cases where subjects had at least moderately varying slopes, wmcorr provided a more visually intuitive estimate of the within-person correlation. Conflicting significance levels or opposite directions of rmcorr and wmcorr served as "warning signs" for potential data quality issues or the need for further data collection. The wmcorr method is proposed as an alternative to the Bland–Altman rmcorr for estimating within-person correlations in repeated-measures data. Researchers are encouraged to estimate both wmcorr and rmcorr, as discrepancies between the two can alert researchers to data patterns that warrant closer inspection before drawing conclusions about within-person relationships.

Keywords: correlation; repeated measures; mixed effects; multilevel

MSC: 62H20

check for updates

Academic Editor: MinJae Lee

Received: 12 December 2024
Revised: 24 January 2025
Accepted: 2 February 2025
Published: 4 February 2025

Citation: Moore, T.M.; Basner, M. An Alternative to the Bland—Altman Repeated-Measures Correlation to Account for Variability of Slopes Across Persons. *Mathematics* **2025**, *13*, 512. https://doi.org/10.3390/math13030512

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

When a data set includes repeated measures, i.e., multiple measurements taken for each person, estimating a correlation between two variables requires separation of variance within and across persons. This is necessary because standard Pearson correlations assume that each observation is independent of all others, but this assumption is violated when multiple measurements come from the same person. When we have repeated measurements from individuals, observations within each person are naturally more similar to each other than to observations from different people. This creates a hierarchical or nested data structure where measurements are clustered within individuals. If we simply calculated a Pearson correlation across all data points while ignoring this nesting, we would conflate two distinct sources of variation: the differences between people's average levels

(between-person variation) and the relationships between variables within each person over time (within-person variation). For example, imagine we are studying the relationship between exercise and mood. Some people might generally exercise more and have better moods overall, creating a positive between-person correlation. However, the relationship we are often really interested in is whether an individual person's mood improves when they exercise more than their personal average—the within-person correlation. By using a repeated-measures correlation, we can specifically examine these within-person relationships while accounting for the fact that each person has their own baseline levels and patterns. This gives us a more accurate understanding of how variables relate to each other at the individual level, rather than mixing this information with broader patterns that exist across different people.

A method for calculating the repeated-measures correlation (r_{rm} or "rmcorr" from here) using a within-person design [1] has gained widespread use, particularly because it (1) provides a single correlation coefficient summarizing all within-person correlations, and (2) remains within the framework of the general linear model, meaning it can be calculated entirely from a table of sums of squared errors (an ANOVA table). To calculate the Bland–Altman rmcorr between X1 and X2, one can simply predict (using regression) X1 using X2 and a series of binary ("dummy") variables representing each subject; for example, if there were 10 subjects, X1 would be predicted by X2 and nine binary variables (with one binary subject variable omitted as the reference category). The partial sums of squares (SS) of this model (usually presented in an ANOVA table) would then be broken down into portions for each independent variable, as well as for the error (residual SS). To find the rmcorr, one would simply divide the SS for the independent variable of interest (X2 in this case) by the sum of that value and the residual SS. The square root of this quotient is the rmcorr. Note the SS for the subjects (binary variables) is ignored, which is by design: the variance attributable to each person's unique mean is considered a "nuisance" in this context. This procedure is quite similar to giving subjects random intercepts in a mixed-effects regression, though there are important differences [2]. The problem with the Bland–Altman approach is that subjects usually do not have the same slope, so estimating random intercepts while assuming the same slope for everyone is often misleading. Of course, reporting a separate correlation for each person defeats the purpose of having a single correlation coefficient to describe the within-person correlation patterns within a sample; therefore, any method summarizing within-person correlations in a single number will have to contend with varying slopes.

Here, we propose an alternative to the Bland–Altman rmcorr. Specifically, we propose calculating all within-person correlations, taking the average of them weighted by the square root of the within-person observations, and testing this weighted average for statistical significance using a one-sample *t*-test also weighted by the square root of the observations. We recommend that researchers use both this method and the original Bland–Altman method for comparison. As we demonstrate below, an extreme inconsistency between these two estimates could suggest methodological or data quality problems that should be addressed before further statistical analyses are conducted.

2. Methods

2.1. Calculating Weighted Mean Within-Person Correlation (wmcorr)

To calculate wmcorr, we first calculate the Pearson correlation for each person in the data set, giving us *N* correlation coefficients. These coefficients are then averaged using the square root of each person's *n* (number of observations) as the weight. For example, assume we had data for three people, with 5, 9, and 15 measurements, respectively. If the correlations for these three people were 0.35, 0.20, and 0.50, respectively, we would take

the weighted average of those three correlations using sqrt(5) = 2.24, sqrt(9) = 3.00, and sqrt(15) = 3.87 as the weights, giving a final wmcorr) of 0.36. Note this weighted mean correlation method is not to be confused with a second Bland–Altman method [3] in the same journal issue, which is for between-subject correlations. They recommend calculating a weighted correlation to determine the between-subject effect, which is unrelated to what we propose here (weighted mean to determine within-subject effect). The square root of n is used as the weight (rather than n) to reflect that the incremental information for estimating the true correlation within a given individual decreases with each additional data point. The p-value for that correlation would be obtained using a weighted one-sample t-test, again using 0.35, 0.20, and 0.50 as the sample values, 0 as the null hypothesis value, and the same set of weights as above. Here, the weighted t value is 3.98 with 2 degrees of freedom, yielding a p-value of 0.06 (non-significant).

A weighted one-sample t-test is the same as the common Student's t-test, except that the weighted mean and weighted variance are used instead of the typical mean and variance (Note that weighted t-tests can also be calculated such that the degrees of freedom are determined by the weights rather than the number of observations, but the present application uses number of observations). The weighted mean is obtained by converting all weights to proportions adding to 1.0, multiplying values by their respective weights, and adding them together. For example, if we had values of −0.60, 0.10, and 0.40, and we had weights of 2, 5, and 10, we would first convert those weights to proportions adding to 1 by dividing them by the total weights. The weight of 2 would become 2/(2 + 5 + 10) = 0.12, 5 would become 5/(2 + 5 + 10) = 0.29, and 10 would become 10/(2 + 5 + 10) = 0.59. The weighted average would therefore be −0.60(0.12) + 0.10(0.29) + 0.40(0.59) = 0.19. The weighted variance is calculated like a typical variance except that the weighted mean is used in finding the squared deviations, among other nuances beyond the present scope. The weighted t-test uses the weighted mean and variance, along with the degrees of freedom determined by the number of persons (not the number of observations). Given the weighted t value and degrees of freedom, the associated p-value is found as though it were a common t value.

An R script for calculating wmcorr is provided in the Supplementary Material, along with the original Bland–Altman data (see below) to use as a demonstration if desired. The functions used in the script include rmcorr() from the *rmcorr* package [4] to calculate the Bland–Altman coefficient, cor() to calculate each within-person correlation, weighted.mean() to calculate the wmcorr, and wtd.t.test() from the *weights* package [5] to calculate the statistical significance of the wmcorr. All four functions are used with default settings. Note that the mean1 = TRUE argument in wtd.t.test() can be changed to mean1 = FALSE if one wishes the degrees of freedom to be calculated using total observations rather than total subjects, but this alternative setting is not tested here.

2.2. Demonstrations

To demonstrate the alternative proposed here, we use it on the real data example presented in the original [1] paper in which pH and $PaCO_2$ (partial pressure of carbon dioxide) share a within-subject relationship. It is unclear why Bland and/or Altman chose the correlation between pH and $PaCO_2$ as the demonstration, but the relationship between pH and $PaCO_2$ is fundamental to understanding how the body maintains the acid–base balance, which is crucial for survival.

We then show an even more extreme example of unusual results obtainable by the Bland–Altman rmcorr method.

2.3. Simulations

Simulations (5000) were run to (1) compare the mean significance levels (p-values) of rmcorr and wmcorr, and (2) determine the relationship between estimated rmcorrs and wmcorrs. Table S1 shows the conditions used, and plain-text details of the conditions shown in this table follow.

Simulation Conditions

Total sample (N). This could vary from 2 to 100, randomly selected for each simulation. The maximum N was set fairly low (100) because preliminary analyses revealed that results in larger samples almost always produced consistent significance decisions (significant or non-significant) between rmcorr and wmcorr.

Number of measurements within each person (n). This could vary from 3 to 40, often varying person-to-person within a single simulation. If the minimum and maximum n were very close (or even the same number) for a given simulation, all simulated subjects would have the same number of measurements; however, if the minimum and maximum were as different as possible (3 and 40, respectively), one subject could have 3 measures while another has 40. A real-world example of this could be a repeated-measures study with 40 visits, where one person showed up for every visit while another person "no showed" for 37 of the visits.

Size of within-person correlations. These could vary from −1.0 to 1.0 and were the main driver (besides N) of the statistical significance of the correlation. If minimum and maximum correlations were very similar (e.g., 0.45 and 0.47, respectively), a plot of all within-person correlations for that simulation would show essentially parallel lines; however, if the minimum and maximum were very different, the plotted slopes could vary wildly.

Scaling of within-person data. This varied from person to person such that one person's repeated measurements could have a standard deviation (SD) of 0.10 while another person's had an SD of 2.00. A real-world example of this could be a study with repeated daily self-reported anxiety and self-reported alcohol use, where the sample included both "normal" and "problem" drinkers.

Absolute shift of within-person data. The X1 and X2 values for each person were shifted randomly up or down by an amount selected randomly from a standard normal distribution. This is equivalent to shifting the intercepts of each person. A real-world example of this could be a repeated-measures study of the relationship between body weight and blood leptin levels, where the sample includes both adults and children. Note that neither the scaling (above) nor the absolute shift manipulation affects the size of the within-person correlation because both are merely linear transformations.

3. Results

3.1. Demonstration Results

Consider the real data example used in the original [1] Bland–Altman paper in which pH and $PaCO_2$ share a within-subject relationship. The figure presented there shows parallel lines fitted to each subject's pH and $PaCO_2$ data (see Figure 1a), representing the relationships estimated and tested by the repeated-measures correlation suggested in that paper. However, Figure 1b shows what those lines would look like if they were fitted to each subject individually, i.e., if the subjects' data were modeled not only with unique intercepts but also unique slopes. When depicted this way, it is less clear that the estimated −0.507 ($p < 0.001$) correlation is correct—indeed, there is reason to doubt that the −0.507 correlation should be considered statistically significant in practice; the p-value associated with it is probably sub-optimal. By contrast, the wmcorr proposed here for this data set is

−0.282 with a *p*-value of 0.256, which is non-significant. The left portion of Table 1 shows the core information used in the wmcorr, namely the eight within-subject correlations, the numbers of points (*n*) used to calculate them, and the square root of the *n* column (the weights used in the weighted average). In this example, the final "warning sign" to researchers would be the conflicting significance levels of rmcorr (highly significant) and wmcorr (non-significant), possibly indicating a problem with data quality, or, at the very least, that more data are necessary before drawing any conclusions about the within-person relationship between $PaCO_2$ and pH.

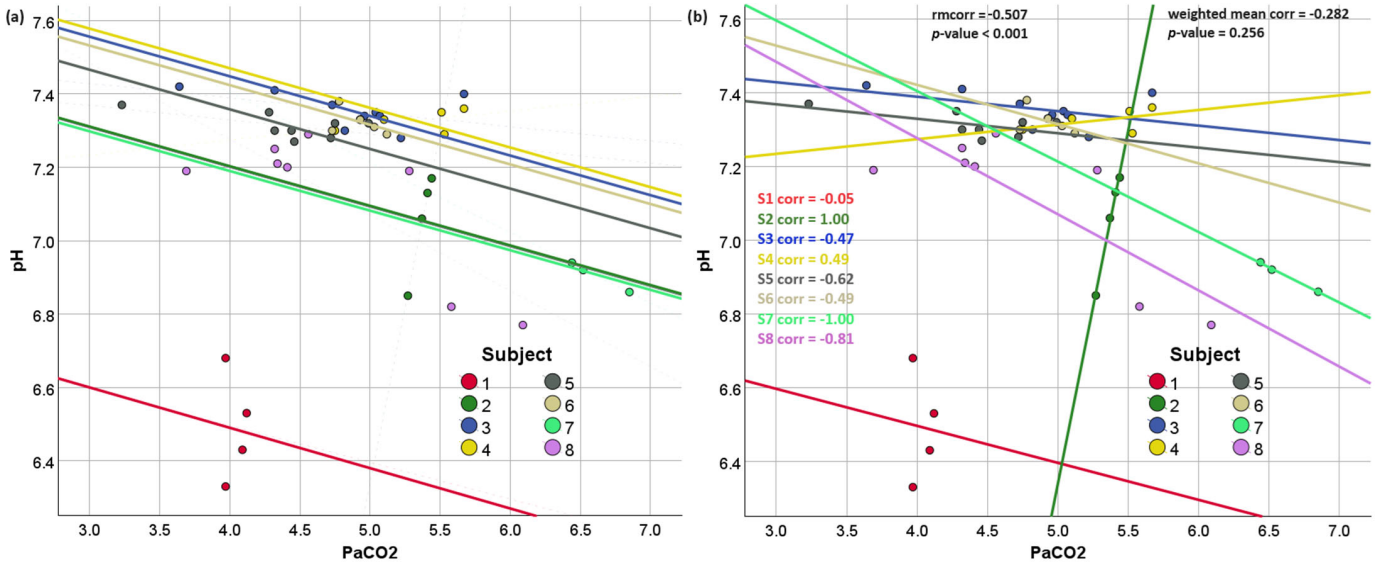


Figure 1. Original Bland–Altman repeated-measures data with parallel slopes (**a**) and varying slopes (**b**).

Table 1. Breakdown of numbers used in calculation of the weighted mean repeated-measures correlation.

Original Bland–Altman [1] Data				Simulated Extreme Example			
Subject	**Within-Subject Corr.**	***n***	**Weight [sqrt(*n*)]**	**Subject**	**Within-Subject Corr.**	***n***	**Weight [sqrt(*n*)]**
1	−0.053	4	2.000	1	−0.909	100	10.000
2	0.997	4	2.000	2	−0.882	100	10.000
3	−0.469	9	3.000	3	−0.928	100	10.000
4	0.494	5	2.236	4	−0.890	100	10.000
5	−0.615	8	2.828	5	−0.905	100	10.000
6	−0.488	6	2.449	6	0.471	100	10.000
7	−0.998	3	1.732				
8	−0.814	8	2.828				

Note. corr = correlation.

Figure 2 shows an even more extreme example of unusual results obtainable by the Bland–Altman rmcorr method. Five of the subjects in Figure 2 have very strong negative correlations ranging from −0.93 to −0.89, while one subject (6) has a moderate positive correlation of 0.47. Due to the wide range of subject 6's data, the overall correlation estimate obtained from the regression and ANOVA table described above is positive (0.242). Further, due to all subjects having roughly the same mean X2 value (~0), the random subject intercepts account for almost no variance. That is, almost all of the explained variance is attributed to the overall positive correlation, resulting in a highly significant *p*-value (2.27×10^{-9}). It is clear that while 0.242 *might* be a satisfactory correlation estimate given for these unusual data, it almost certainly should not be considered statistically significant with such high confidence, given five out of the six subjects show very strong correlations

in the opposite direction. The wmcorr for these data is heavily influenced by the five within-person correlations around −0.90, yielding a wmcorr of −0.67 with a significant *p*-value (0.032). This example is useful for demonstrating the core differences between rmcorr and wmcorr. While rmcorr is focused on control for variance due to subjects differing in mean levels (i.e., between-subject variance), rmcorr ignores mean levels entirely, focusing on each subject's correlation "in a vacuum" and then aggregating all subjects' correlations (weighted).

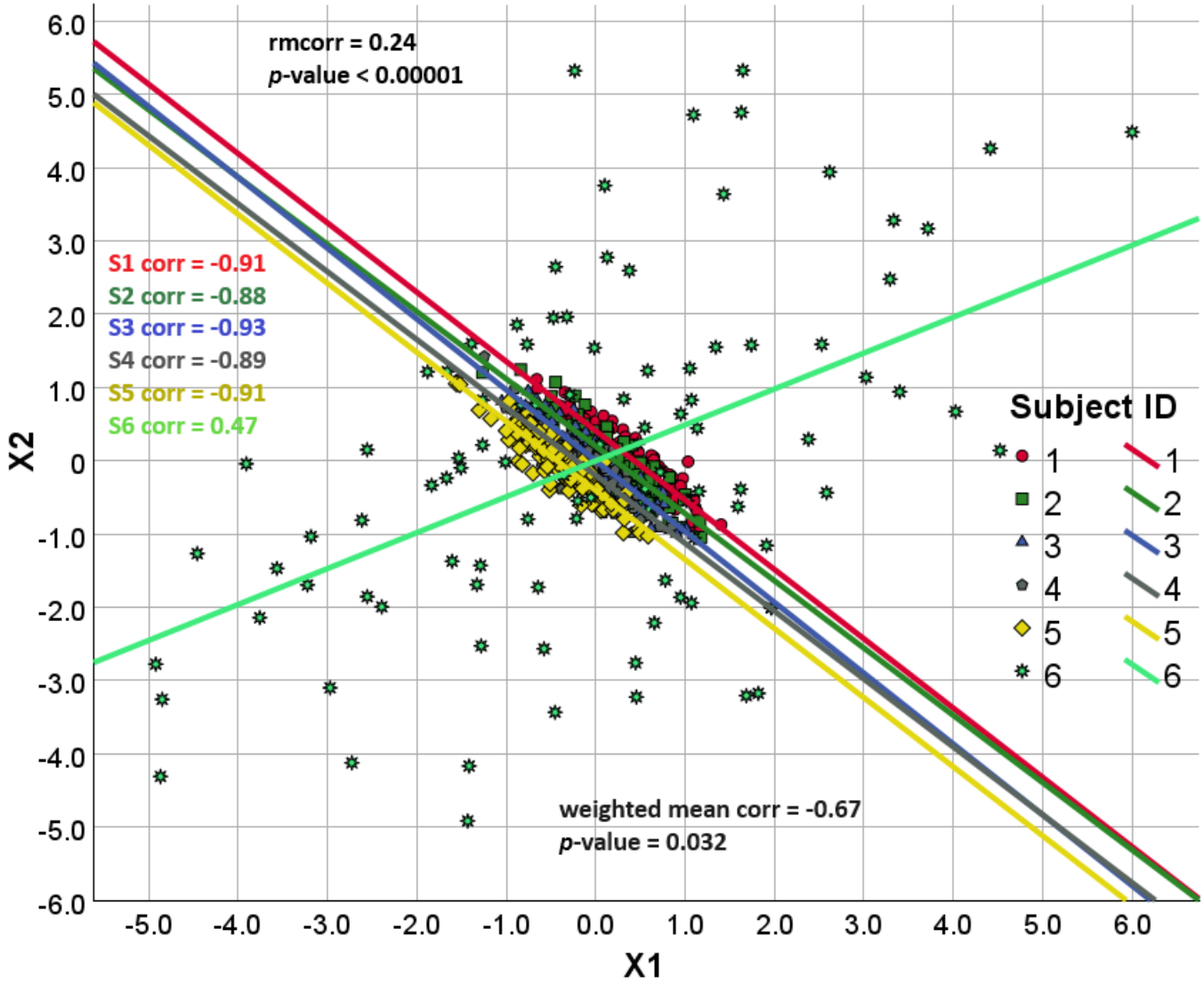


Figure 2. Extreme example of high-confidence Bland–Altman correlation contradicted by weighted mean correlation.

The right portion of Table 1 shows the core information used for wmcorr; for simplicity, all subjects were simulated with the same *n*. In this example (Figure 2), the "warning signs" are that (1) levels of significance of rmcorr and wmcorr are quite different (though both are significant), and, critically, (2) the estimates are in opposite directions. This is another example where thorough inspection of the data is warranted.

3.2. Simulation Results

Supplementary Figure S1 shows how the original Bland–Altman rmcorr *p*-values compare to the weighted mean (wmcorr) *p*-values, separated by size of rmcorr (by tertile) and *N* (median split). Based on the original rmcorr *p*-values (gray bars), almost any rmcorr > 0.13 would be considered significant as long as *N* exceeds 50 (bottom row-panel of Figure S1). By contrast, using the new *p*-values (orange bars), there is a steadier decline in the average *p*-value up to ~0.32.

Figure 3 shows four simulated example data sets ranging from $N = 2$ to $N = 5$ with varying combinations of rmcorr and wmcorr results. Panel "a" shows data for two persons with opposite relationships between X1 and X2, seeming to "cancel each other out" for an overall relationship near zero. Notably, the estimated Bland–Altman rmcorr is 0.50 with a highly significant $p < 0.001$, but wmcorr (0.073) falls closer to what would be expected given the visual. Panel "b" ($N = 3$) shows an example where rmcorr reveals a significant positive relationship (0.447); however, wmcorr, while being larger in size (0.568) than the rmcorr, is found to be non-significant ($p = 0.154$), likely due to the small *N*. Panel "c" shows an example analogous to that shown in Figure 2 in the main text: due to the large variance of one subject, rmcorr indicates a substantial correlation (−0.672) in the opposite direction to the majority of subjects, and wmcorr indicates as much. Panel "d" ($N = 5$) shows an example of both measurements agreeing, where rmcorr is positive (0.616) and significant, and wmcorr is significant ($p = 0.035$) and similar in magnitude to rmcorr. Finally, Figure 4 shows the relationship between the Bland–Altman estimated rmcorr and the weighted average estimate (wmcorr). These estimates correlate at 0.98, indicating that in the vast majority of cases, the wmcorr and rmcorr will closely agree. To confirm absolute (rather than only relative) agreement between rmcorr and wmcorr, we estimated the intraclass correlation coefficient (ICC) as well, and indeed, the value remains high (0.98).

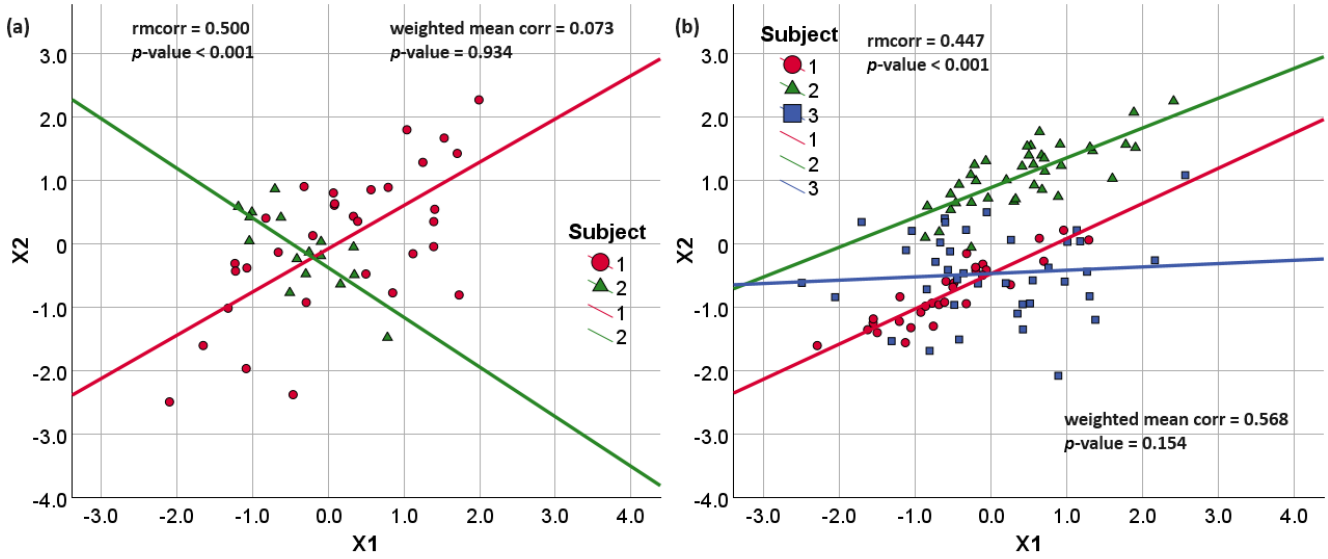


Figure 3. *Cont.*

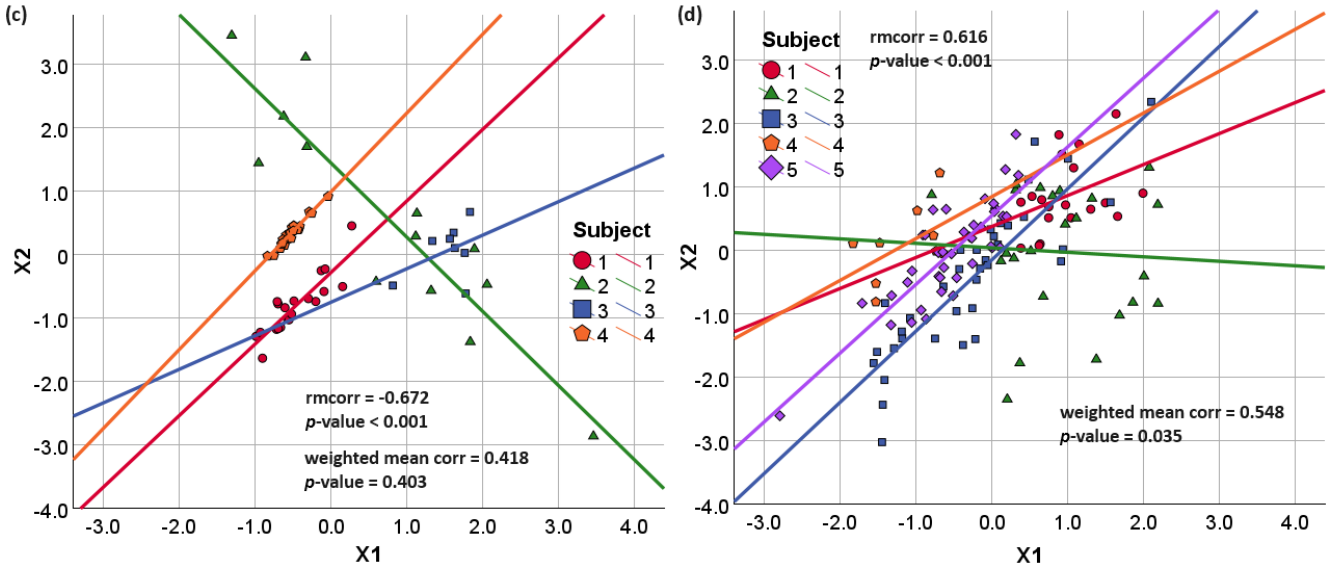


Figure 3. Simulated examples showing varying combinations of significance levels across correlation and *p*-value types using two subjects (**a**), three subjects (**b**), four subjects (**c**), and five subjects (**d**).

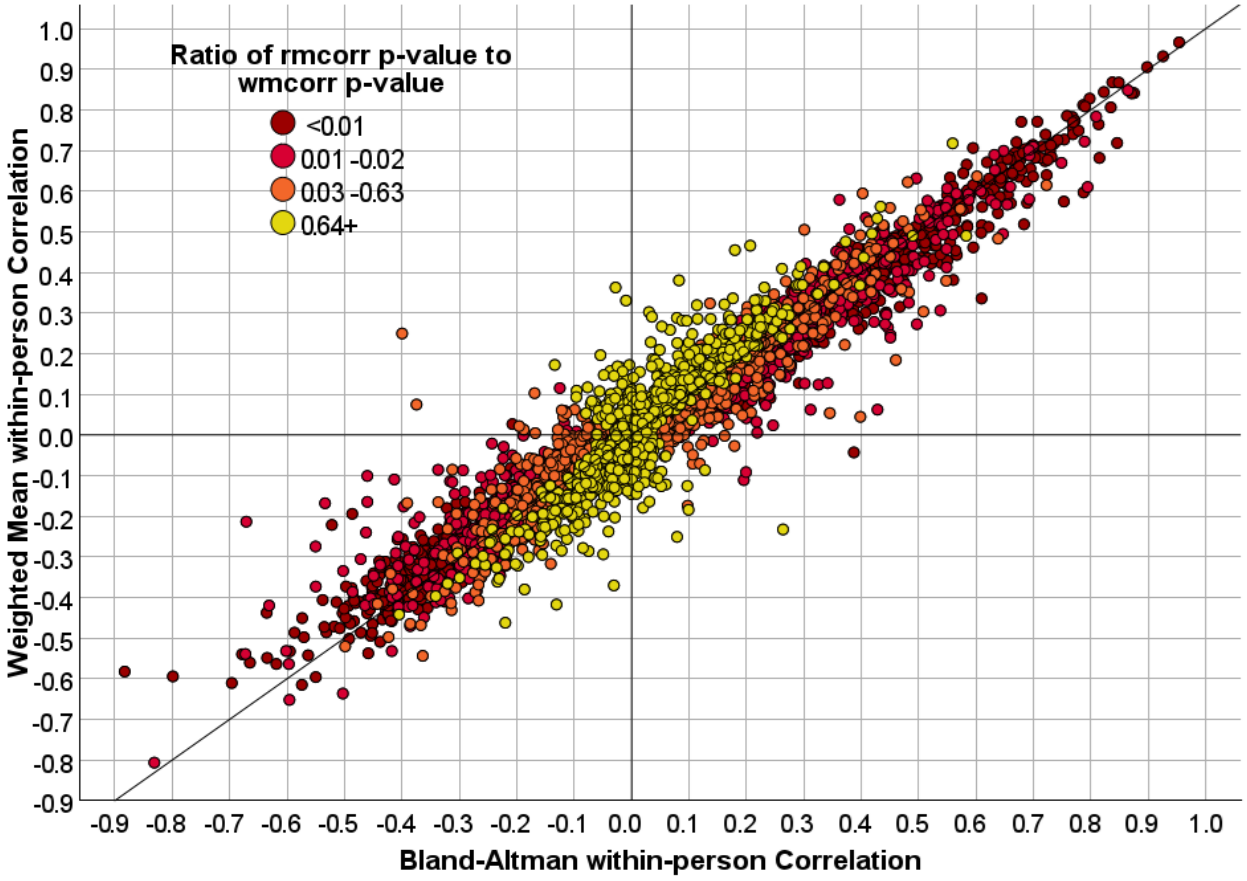


Figure 4. Relationship between Bland–Altman and weighted mean correlations, separated by influence of the multiplier.

Regarding the statistical properties of the wmcorr estimator, we can say based on Supplementary Figure S1 that wmcorr tends to be less powerful than rmcorr, meaning Type I errors are less likely with wmcorr. Regarding bias, this was tested by conducting simulations identical to those described above except that the within-person correlation was held constant across individuals. That is, with each use of the simCor() function

(from the *psych* package [6]) to simulate data for a specific person, the value used was identical to that of all other individuals in that simulation; this produced a "true" repeated-measures correlation that can be compared with the estimated correlations (rmcorr and wmcorr). Supplementary Figure S2 shows the distributions of errors for these simulations by N. The symmetries of both distributions (means ≈ 0) suggest absence of bias, and as expected, at lower Ns, errors tended to be more variable (yellow, wider distribution) than at higher Ns. The absence of bias was confirmed via scatterplots of true absolute correlations (x-axis) and errors (y-axis), in which the relationship was almost exactly zero for both rmcorr and wmcorr. However, one interesting result that stands out from Figure S2 is the larger variability of errors for rmcorr (SD = 0.108) than for wmcorr (SD = 0.086), suggesting generally higher accuracy for wmcorr. To investigate this, we compared mean absolute errors for rmcorr and wmcorr at varying levels of N and true absolute correlation. Supplementary Figure S3 shows these comparisons, and as is apparent from the top four and leftmost three panels, rmcorr is significantly less accurate than wmcorr when N *or* the true absolute correlation is low. We do not provide p-values for these results because they are arbitrarily dependent on the number of simulations, but the pattern is clear from the results in Figure S3 that the accuracy of wmcorr is superior under some conditions.

To summarize the simulation results, they demonstrate clearly that (1) the Bland–Altman rmcorr and weighted average wmcorr will be very similar in most cases; (2) unusual cases like the one demonstrated in Figure 2 will tend to be "caught" by wmcorr; (3) neither method showed bias in estimating the true correlations, though rmcorr exhibited more variable errors (SD = 0.108) compared to wmcorr (SD = 0.086); and (4) wmcorr has better statistical properties in certain conditions, showing higher accuracy when sample sizes are small or true correlations are low.

4. Discussion

We propose an alternative to the Bland–Altman repeated-measures correlation (rmcorr) for estimating within-person correlations between two variables in a repeated-measures data set. Our proposed method, the weighted mean within-person correlation (wmcorr), calculates the average of all within-person correlations, weighted by the square root of the number of observations for each person. We demonstrate that in most cases, rmcorr and wmcorr will yield similar results; however, in cases where subjects have at least moderately varying slopes, wmcorr may provide a more visually intuitive estimate of the within-person correlation. Further, simulation results showed that while wmcorr had less statistical power than rmcorr (reducing Type I errors), neither method exhibited systematic bias in estimating correlations; however, wmcorr demonstrated superior accuracy, particularly when sample sizes or true correlations were low, as evidenced by its smaller error variability (SD = 0.086 vs. 0.108). While we call wmcorr an "alternative" because it could be used as a standalone method, we encourage researchers to estimate both wmcorr and the original rmcorr. Conflicting significance levels or opposite directions of rmcorr and wmcorr can serve as "warning signs" for researchers, indicating potential data quality issues or the need for further data collection before drawing conclusions about within-person relationships.

The method proposed here shares some fundamental characteristics with mixed-effects regression models, but differs in important technical aspects. Both methods aim to account for the nested structure of repeated-measures data and focus on within-person relationships, although mixed-effects models also provide between-subject effects. The weighting by square root of sample size in the proposed approach parallels how mixed models naturally give more weight to subjects with more observations, as these subjects provide more reliable information about within-person patterns. However, the methods diverge in their underlying "machinery" and assumptions. Mixed-effects regression explicitly models

both fixed and random effects, allowing for variation in both intercepts and slopes across individuals, while simultaneously estimating an overall population-level relationship. It achieves this through sophisticated maximum likelihood estimation that considers the entire data structure. Our method, by contrast, takes a two-stage approach, first calculating individual correlations, and then combining them through weighted averaging. This makes our method more computationally straightforward and potentially more intuitive to researchers, because it is clear how each person's data contributes to the final estimate. It also avoids some of the complex assumptions about the distributions of random effects that mixed models require. However, this simplicity means our method might be less efficient at using all available information in the data, particularly when some subjects have very few observations or when there are missing data patterns that mixed models could handle more elegantly through their likelihood-based framework.

The method proposed here is motivated primarily by potential real-world applications of the Bland–Altman rmcorr. From a scientific perspective, the *p*-value obtained from the original method comes from a known distribution of a test statistic given some degrees of freedom, so it will always be "correct" in that sense, i.e., the probability of obtaining the rmcorr result by chance would be the *p*-value already offered by the method presented in the original Bland–Altman paper [1]. However, because statistical significance is often used as a threshold to indicate whether a model should be applied in the real world, we hope we have demonstrated how some significant rmcorr models could be applied unreasonably. Using the original data as an example, the significant unadjusted *p*-value might lead some to conclude that if someone's $PaCO_2$ levels changed by one standardized unit, his/her pH would decrease by 0.507 standardized units. While this might be true on average, the results in Figure 1 demonstrate that the variability in expected pH is enormous; indeed, two of the eight subjects show an *increase* in pH, one quite substantially. Our alternative suggestion, wmcorr, provides a reasonable alternative to the Bland–Altman [1] method. Our hope is that cases in which the original and weighted average methods provide quite different results will alert researchers to data patterns that warrant closer inspection.

While wmcorr is a viable addition to the "correlation toolbox", several limitations should be considered. First, our simulation conditions, though varied, do not capture all possible real-world data scenarios, particularly those with extreme outliers or highly non-linear relationships. Second, the method's reliance on calculating individual correlations means it requires at least three observations per person to compute a correlation coefficient, potentially limiting its application in studies with sparse sampling or irregular measurement intervals. Third, the square root weighting approach, while theoretically justified, is just one possible weighting method; alternative schemes might be more appropriate in certain contexts but were not explored in this study. Fourth, while the examples used in the present study are compelling, they are likely to be quite rare in practice. Indeed, something as extreme as that shown in Figure 2 would likely only occur if there were serious problems in data entry or cleaning. However, this is also one of the strengths of the method: it will detect such problems when they might otherwise go undetected. Finally, while our method can identify potentially problematic data patterns, it does not provide specific guidance on how to address such issues beyond suggesting further data collection or inspection.

5. Conclusions

The introduction of the weighted mean within-person correlation (wmcorr) is a methodological advancement in analyzing repeated-measures data, offering researchers a complementary tool alongside the established Bland–Altman rmcorr method. While both approaches aim to capture within-person relationships in nested data structures, wmcorr's intuitive weighting system and ability to detect potentially problematic data patterns make

it particularly valuable for research applications. Our simulations demonstrate that these methods typically converge on similar results, providing mutual validation when they align. However, in cases where they diverge, this disagreement serves as a crucial diagnostic tool, alerting researchers to potential data quality issues or the need for additional data collection. This "dual-method" approach ultimately strengthens the robustness of repeated-measures analysis and helps ensure that statistical significance translates meaningfully to real-world applications, particularly in fields like medicine and psychology where individual differences can have profound practical implications.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math13030512/s1, Figure S1: Comparison of Original (rmcorr) and Weighted Mean (wmcorr) *p*-values by N and Size of rmcorr; Figure S2: Error distributions for Original (rmcorr) and Weighted Mean (wmcorr) correlations, by N; Figure S3: Mean absolute error (and 95% Confidence Intervals) for rmcorr and wmcorr, by N and absolute size of true (population) correlation; Table S1: Simulation Conditions.

Author Contributions: Conceptualization, T.M.M. and M.B.; methodology, T.M.M. and M.B.; software, T.M.M.; validation, T.M.M. and M.B.; formal analysis, T.M.M.; investigation, T.M.M. and M.B.; resources, T.M.M. and M.B.; data curation, T.M.M.; writing—original draft preparation, T.M.M. and M.B.; writing—review and editing, T.M.M. and M.B.; visualization, T.M.M.; supervision, T.M.M. and M.B.; project administration, T.M.M. and M.B.; funding acquisition, T.M.M. and M.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by NIMH grants MH117014 and MH119219 and NASA grants 80NSSC19K1046 and 80NSSC21K1698.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bland, J.M.; Altman, D.G. Calculating correlation coefficients with repeated observations: Part 1—Correlation within subjects. *BMJ* **1995**, *310*, 446. [CrossRef]
2. Bell, A.; Fairbrother, M.; Jones, K. Fixed and random effects models: Making an informed choice. *Qual. Quant.* **2019**, *53*, 1051–1074. [CrossRef]
3. Bland, J.M.; Altman, D.G. Calculating correlation coefficients with repeated observations: Part 2—Correlation between subjects. *BMJ* **1995**, *310*, 633. [CrossRef] [PubMed]
4. Bakdash, J.; Marusich, L. rmcorr: Repeated Measures Correlation. R Package Version 0.6.0. 2023. Available online: https://CRAN.R-project.org/package=rmcorr (accessed on 15 January 2024).
5. Pasek, J.; Tahk, A.; Culter, G.; Schwemmle, M. Weights: Weighting and Weighted Statistics. R Package Version 1.0.4. 2018. Available online: https://CRAN.R-project.org/package=weights (accessed on 15 January 2024).
6. Revelle, W. psych: Procedures for Psychological, Psychometric, and Personality Research. R Package Version 2.4.2. 2024. Available online: https://CRAN.R-project.org/package=psych (accessed on 15 January 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.