UC Irvine
UC Irvine Previously Published Works

Title

The case for the repeatability intra-class correlation as a metric of precision for salivary bioscience data: Justification, assessment, application, and implications.

Permalink

https://escholarship.org/uc/item/80v9z7w2

Authors

Riis, Jenna L
Ahmadi, Hedyeh
Hamilton, Katrina R
et al.

Publication Date

2021-06-01

DOI

10.1016/j.psyneuen.2021.105203

Peer reviewed

The case for the repeatability intra-class correlation as a metric of precision for salivary bioscience data: Justification, assessment, application, and implications

Jenna L. Riis[a,b,*], **Hedyeh Ahmadi**[a], **Katrina R. Hamilton**[a], **Crystal I. Bryce**[c], **Clancy Blair**[d], **Douglas A. Granger**[a,b,e,f,g]

[a]Institute for Interdisciplinary Salivary Bioscience Research, University of California, Irvine, CA, USA

[b]Department of Psychological Science, University of California, Irvine, CA, USA

[c]T. Denny Sanford School of Social & Family Dynamics, Arizona State University, Tempe, AZ, USA

[d]Department of Population Health and Department of Applied Psychology, New York University, New York, NY, USA

[e]Department of Acute and Chronic Care, Johns Hopkins University School of Nursing, Baltimore, MD, USA

[f]Department of Pediatrics Johns Hopkins University School of Medicine, Baltimore, MD, USA

[g]Salivary Bioscience Laboratory and Department of Psychology, University of Nebraska, Lincoln, NE, USA

Abstract

Best practice standards for measuring analyte levels in saliva recommend that all biospecimens be tested in replicate with mean concentrations used in statistical analyses. This approach prioritizes minimizing laboratory-based measurement error but, in the process, expends considerable resources. We explore the possibility that, due to advances in salivary assay precision, the contribution of laboratory-based measurement error in salivary analyte data is very small relative to more important and meaningful variability in analyte levels across biological replicates (i.e., between different specimens). To evaluate this possibility, we examine the utility of the repeatability intra-class correlation (rICC) as an additional index of salivary analyte data precision. Using randomly selected subsamples (*N*s=200 and 60) of salivary analyte data collected as part of a larger epidemiologic study, we compute the rICCs for seven commonly assayed salivary

---

[*]Correspondence to: Institute for Interdisciplinary Salivary Bioscience Research, University of California at Irvine, 4201 Social and Behavioral Sciences Gateway, Irvine, CA 92697-7085, USA. jriis@uci.edu (J.L. Riis).

measures in biobehavioral research - cortisol, alpha-amylase, c-reactive protein, interlekin-6, uric acid, secretory immunoglobulin A, and testosterone. We assess the sensitivity of rICC estimates to assay type and the unique distributions of the underlying analyte data. We also use simulations to examine the bias, precision, and coverage probability of rICC estimates calculated for small to large sample sizes. For each analyte, the rICCs revealed that less than 5% of variation in analyte levels was attributable to laboratory-based measurement error. rICC estimates were similar across all analytes despite differences in analyte levels, average intra-assay coefficients of variation, and in the distributional properties of the data. Guidelines for calculating rICC are provided to enable investigators and laboratory staff to apply this metric and more accurately quantify, and communicate, the magnitude of laboratory-based measurement error in their data. By helping investigators scale measurement error relative to more scientifically meaningful variability between biological replicates, the application of the rICC has the potential to influence research strategies and tactics such that resources (e.g., finances, effort, number/volume of biospecimens) are allocated more efficiently and effectively.

Keywords

Salivary bioscience; Repeatability intra-class correlation; Technical replicate; Measurement error; Intra-assay coefficient of variation

1. Introduction

The past few decades have witnessed a macro-level trend that involves the widespread integration of biological measures in research studies across scientific disciplines. The ease of use and minimally invasive nature of saliva as a biospecimen has been especially valued in this endeavor as it enables the study of complex models of behavior, cognition, and health in laboratory, quasi-naturalistic, and real-world settings. There has been significant progress in developing and refining saliva collection and measurement methods, as well as expanding the number of salivary analytes involved in this research (see (Granger and Taylor, 2020) for review). Laboratory assay protocols for salivary biomeasures have evolved from in-house modifications of serum-based assays to modern methods specifically designed for saliva. These state-of-the-art assays are now widely commercially available, and, in many cases, designed to satisfy the immunodiagnostic industry's rigorous criteria for *precision* and *reproducibility* (see (Granger and Gaitonde, 2020)).

Advances in salivary analyte measurement, however, have not necessarily been matched with comparable adjustments to research laboratory operating procedures. Current best practice standards in salivary analyte measurement generally require all biospecimens be tested in replicate. This approach prioritizes, and expends considerable resources to minimize, laboratory-based measurement error- a source of variability in analyte data that is generally very small. Further, this focus on measurement error at the level of individual analyte determinations is often misaligned with both our overarching research goals and the sophisticated statistical strategies we employ (see (Riis et al., 2020) for review).

We raise the possibility that, in some circumstances, inclusion of more data points per participant (i.e., biological replicates), or a larger sample size, or both, would be priority

over expending resources to test biospecimens in replicate (i.e., technical replicates). To explore this possibility, and objectively define these circumstances, we examine the repeatability intra-class correlation (rICC) as an index of biomeasure precision and data quality. We demonstrate the use and interpretation of the rICC in salivary bioscience and provide guidance for future researchers to use this index to inform their laboratory procedures, study designs, and resource allocation decisions.

1.1. Conventional metric of biomeasure precision: limitations of the intra-assay coefficient of variation (CV)

Currently, salivary bioscientists largely rely on the intra-assay CV as a measure of precision. The intra-assay CV is a point estimate reflecting variance in measured analyte levels for individual biospecimens tested in replicate. It is calculated by computing the relative variance across replicate determinations using the formula: $\left(\frac{SD(replicate_1, replicate_n)}{mean(replicate_1, replicate_n)} \right) \times 100$.

The intra-assay CV is typically expressed as a percentage with perfect agreement indicated by an intra-assay CV of 0%. In general, immunodiagnostic industry standards expect an average intra-assay CV across all study samples 5% (Chard, 1990). The intra-assay CV has well-known limitations. It is percentage-based, and its magnitude depends on the level of the analyte. That is, the same variance across replicate determinations yields a small CV% when analyte concentrations are high and a large CV% when analyte concentrations are low. Also, as a measure of precision at the level of the individual biospecimen, the intra-assay CV does not relate variance across replicate determinations to the total variance in analyte levels across the study, thereby hindering the consideration of replicate variance within the larger context of the study.

1.2. The repeatability intra-class correlation

The rICC is an additional approach to measuring salivary biomeasure precision. Within the context of biospecimen testing, the rICC provides information about the proportion of total variation in analyte concentrations across a study that is attributable to variation across repeated measurements of the same biospecimen (Lessells and Boag, 1987; Shrout and Fleiss, 1979). The rICC is measured on a scale of 0–1 with perfect precision indicated by 1 (meaning all the variation in analyte measurements in the study is due to differences in analyte levels across biospecimens and there is no laboratory-based measurement error). The rICC is estimated using the multilevel structure of the analyte data (replicates nested within biospecimen), and this estimation allows for the calculation of confidence intervals (CIs) around the rICC.

The rICC addresses two of the primary limitations of the intra-assay CV. First, by expressing measurement error in salivary determinations as a proportion of total variance in analyte levels across the study, rather than at the level of the individual biospecimen, the rICC allows the researcher to contextualize the relative importance of measurement error. Also, the estimation of CIs around the rICC provides a more detailed understanding of expected measurement error. When calculated during the early stages of a study, rICC CIs can objectively inform study planning for the number of technical replicates needed – more duplicate testing if measurement error is meaningful, or more singlet testing if measurement

error is negligible. These decisions have important implications for resource allocation (Table 1). For example, if measurement error is minimal and singlet testing is sufficient, resources (e.g., time, funds, reagents, lab supplies) otherwise reserved for and consumed by technical replicates could be used to assess additional participants, time points, and/or analytes.

1.3. Present study

This study examines the utility and value of the rICC as a metric for the salivary bioscience community, provides guidelines for its calculation and interpretation, and highlights its use in the effort to realign advances in the precision of modern salivary assays with salivary bioscience research design and resource allocation priorities. We explore the use of the rICC as an index of biomeasure precision for seven salivary analytes commonly employed in biobehavioral research: cortisol, alpha-amylase (sAA), c-reactive protein (CRP), interlekin-6 (IL-6), uric acid (UA), secretory immunoglobulin A (SIgA), and testosterone. We compare rICCs across different assay types (singleplex vs. multiplex; immunoassay vs. kinetic reaction), examine the sensitivity of the rICC to data distribution issues common in salivary bioscience (i.e., skewed distributions and extreme/influential data points), and use simulated data to assess the stability of rICC estimates calculated using datasets of various sizes.

2. Material and methods

This study uses archival salivary analyte data assayed as part of a larger epidemiologic study - the Family Life Project (FLP) 12-year follow-up assessment. Descriptions of the FLP study (https://flp.fpg.unc.edu/) have been presented elsewhere (Vernon-Feagans and Cox, 2013). rICC analyses for all salivary analytes, except sAA, used data collected from a single, baseline saliva sample from 200 FLP participants (45% female; 44% African American). For sAA analyses, approximately 10% of saliva samples in the overall study were tested in duplicate, and all sets of complete duplicate determinations of sAA from baseline saliva samples were used in our analyses ($N = 60$; 38% female; 30% African American).

2.1. Collection of samples and determination of salivary analytes

Unstimulated whole saliva samples were self-collected by participants at home. Samples were immediately frozen, then transported frozen to the Institute for Interdisciplinary Salivary Bioscience Research (IISBR) laboratory at the University of California, Irvine where they were stored at $-80\,°C$. IISBR laboratory technicians operate under Good Laboratory Practice (GLP) guidelines. Liquid handling equipment, plate washers, and plate readers are maintained and calibrated following Standard Operating Procedures. On the day of assay, samples were thawed and centrifuged to remove mucins. All biospecimens were assayed in duplicate, and the average inter-assay CVs were less than 15%.

2.1.1. Cortisol—As described (see (Blair et al., 2011)), samples were assayed for salivary cortisol using commercially available, competitive immunoassay ELISA kits (Catalog #1-3002, Salimetrics, Carlsbad, CA). The test volume was 25 μL, and the assay limit of sensitivity was < 0.007 ug/dL with a measurement range up to 3.00 ug/dL.

2.1.2. Alpha-amylase—As described (see (Granger et al., 2007)), samples were assayed for sAA using commercially available, kinetic enzymatic kits (Catalog #1-1902, Salimetrics, Carlsbad, CA). The test volume was 10 μL, and the sensitivity range for this assay was 0.4–400 U/mL.

2.1.3. C-reactive protein—Samples were assayed for CRP using the Human CRP (Vascular Injury Panel 2) V-Plex kits (Ref# K0080900, MSD) following the manufacturer's guidelines. CRP concentrations (pg/mL) were determined with MSD Discovery Workbench Software (v. 4.0) using curve fit models (4-PL with a weighting function option of $1/y2$). Samples were diluted 5-fold in MSD Assay Diluent 101 with a test volume is 20 μL. The sensitivity range for this assay was 9.85–1010000 pg/mL.

2.1.4. Interleukin-6—Samples were assayed for IL-6 using a V-Plex Human Proinflammatory cytokine Panel (4-plex) manufactured by Meso Scale Discovery (MSD, Gaithersburg, MD, Ref# K008074) (Riis et al., 2014). A testing volume of 25 μL per sample was diluted 2-fold with MSD Assay Diluent 2 before being added to the plate for testing. The 4-plex Multi-Spot Array assay was run following the manufacturer's recommended protocol without modification. Cytokine concentrations were determined with MSD Discovery Workbench Software (v. 4.0) using curve fit models (4-PL with a weighting function option of $1/y^2$). The sensitivity range for this assay was 0.38–1530 pg/mL.

2.1.5. Uric acid—Following Riis et al. (2018), samples were assayed for UA using enzymatic assay kits (Catalog #1-3802, Salimetrics, Carlsbad, CA). The test volume was 10 μL, and the sensitivity for this assay was 0.07 mg/dL with a measurement range up to 20 mg/dL.

2.1.6. Secretory IgA—As described (see (Laurent et al., 2015)), samples were assayed for salivary SIgA using commercially available, indirect competitive enzyme immunoassay kits (Catalog # 1–1602, Salimetrics, Carlsbad, CA). The manufacturer's protocol was followed without modification. The test volume was 10 μL, and samples were tested at a 1:5 dilution. The sensitivity for this assay ranged from 12.5 to 3000 μg/mL. SIgA concentrations were calculated from a standard curve generated using a 4-parameter non-linear regression curve fit (Gen5, BioTek, Winooski, VT).

2.1.7. Testosterone—As described (see (Rodriguez et al., 2020)), samples were assayed for testosterone using commercially available, enzyme-linked immunosorbent assay kits (Catalog # 1-2402, Salimetrics, Carlsbad, CA) following the manufacturer's protocol. The test volume was 25 μL, and the sensitivity for this assay ranged from 1 to 600 pg/mL.

2.1.8. Salivary assay data quality assurance and quality control—Raw salivary analyte data were used as measured by the first round of assay testing. Determinations that would be flagged as requiring a repeat test during conventional quality assurance/quality control (QA/QC) processes at the first round (e.g., due to high intra-assay CVs) were included in the analyses to protect the rICC calculations against artificial inflation due to laboratory QA/QC processes.

2.2. Analytic plan

2.2.1. Preliminary analyses—For each analyte, non-detect determinations (i.e., cases with analyte concentrations that fell outside the range of assay measurement) were removed prior to the random sampling of 200 individuals. The distribution and range of salivary analyte data were examined using descriptive statistics (Table 2), and the impact of transforming the data on the skew and kurtosis of the data was examined.

2.2.2. Repeatability intra-class correlations—The rICC was computed using the equation $\frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}$ where σ_α^2 is the between-individual variance and σ_ϵ^2 is the within-individual variance (Nakagawa and Schielzeth, 2010). We used a linear mixed modeling-based estimation of the rICC with two levels. Duplicate analyte determinations (level 1) were nested within participant (level 2; details provided in the Supplementary Materials, Methods section). CIs were estimated using a parametric bootstrap (see (Stoffel et al., 2019) for more details). We used restricted maximum likelihood estimation to minimize bias in the estimated variance components and examined the normality, homoscedasticity, and outliers of the level-1 and level-2 residuals to assess model fit.

2.2.3. Repeatability intra-class correlations and assay type, non-normal data, and influential data points—We compared rICC estimates across analyte to assess the sensitivity of the rICC to assay type and quality. For each analyte, rICC analyses were conducted using both the raw, untransformed salivary data as well as data that were transformed to improve the normality of the distributions. We then compared rICC estimates within analyte to assess the impact of non-normal data distributions on the rICC. Also, for each rICC analysis conducted with log-transformed analyte data, we identified potentially influential data points using Q–Q plots and Cook's distance criteria. These data points were then excluded from the analytic dataset, creating a "trimmed" dataset. rICC estimates were recalculated using these trimmed datasets. Estimates generated using complete and trimmed datasets were compared within analyte to assess the impact of potentially influential data points on estimated rICC parameters.

2.2.4. Repeatability intra-class correlations and simulated sample size—
Salivary biomeasure data with two replicate determinations per participant were simulated using a two-level, linear mixed model shown in the Supplementary Materials, Methods section. Simulations were created by specifying between-individual (i.e., σ_a^2) and within-individual (i.e., σ_ϵ^2) variances which, in turn, would define the rICC. Three combinations of variance parameters were used to create rICCs of 0.90., 0.95, and 0.99: 1) $\sigma_a^2 = 0.99$ & $\sigma_\epsilon^2 = 0.11 \Rightarrow ICC = 0.90$; 2) $\sigma_a^2 = 0.58$ & $\sigma_\epsilon^2 = 0.03 \Rightarrow ICC = 0.95$; 3) $\sigma_a^2 = 0.99$ & $\sigma_\epsilon^2 = 0.01 \Rightarrow ICC = 0.99$. Samples sizes of 50, 100, 150, and 200 were simulated for each rICC calculation. Sample sizes of 250, 300, and 350 were also simulated for rICCs of 0.90 and 0.95. These analyses allowed us to assess the estimated variability around the "true" rICC that is expected when estimates are calculated using datasets of various sample sizes. Five-hundred simulations were created, and bootstrapped CIs were estimated (with 1000 bootstrap iterations).

We adopted a preliminary lower-limit threshold of 0.95 for the rICC (Harper, 1994) and evaluated estimated rICC CIs using this criterion. For a fixed rICC, the sample size at which the lower bound of the rICC CI was approximately   0.95 was identified as the smallest sample size that would support a reliably high rICC estimate using this preliminary criterion.

All analyses were conducted in R (R Core Team, 2019). rICC estimations and model checks were conducted using the rptR (Stoffel et al., 2019) and lme4 (Bates et al., 2020) packages, respectively. Our annotated R script is provided in the Supplementary Materials.

3.1. Preliminary analyses

Descriptive statistics and intra-assay CVs for all analytes are presented in Tables 2 and 3. Log-transforming the analyte data mostly improved the normality of the distributions for all analytes (Table 2). The average intra-assay CVs ranged from 1.91% to 8.00%, and the number of cases with intra-assay CVs exceeding 5% and 15% ranged from 10 to 105 (5.50–52.50%) and from 0 to 24 (0.00–12.00%), respectively, across all analytes (Table 3).

3.2. Repeatability intra-class correlations and assay type, non-normal data, and influential data points

All analyte rICCs were very high (ranging from 0.971 to 1.000 across all analytes and conditions; Table 3). For all analytes, rICC point estimates were similar when comparing results from models using the raw vs. log-transformed data (Table 3). When rICC analyses excluded potentially influential points (6–22 cases excluded; Table 3), the rICC CIs overlapped for all analytes, except for CRP, UA, and sAA (using log-transformed data; Table 3). Although not overlapping, the rICCs for CRP, UA, and sAA using the untrimmed and trimmed datasets were very precise with narrow CIs (Table 3).

3.3. Repeatability intra-class correlations and simulated sample size

For all simulations, the difference between the "true rICC" and the mean of estimated rICCs was   0.001. The coverage probability for all the sample sizes tested was   92%, meaning that our simulated CIs covered the true rICC value at least 92% of the time.

When the true rICC of the data was 0.99, our findings show that we can estimate reliably high rICCs (assuming a preliminary lower bound threshold of   0.95) using any sample size from 50 to 200 biospecimens (Fig. 1). For data with a true rICC of 0.95, however, the lower bound of the CI was as low as 0.92 when estimated with a sample size of 50 and as high as 0.94 when estimated with a sample size of 200 (Fig. 1). With a true rICC of 0.90, estimation was relatively variable, ranging from 0.84 to 0.96 when calculated using 50 biospecimens and the range only improved to 0.87–0.93 when calculated using a sample size of 200 (Fig. 1). Estimation of the rICCs with sample sizes > 200 showed limited added benefits in terms of precision, bias, and coverage probability (Supplementary Materials, Results section and Supplemental Fig. 1).

4. Discussion

Our findings demonstrate the utility of the rICC in evaluating the precision of salivary analyte data. Among the advantages of using the rICC in salivary bioscience studies is its easy translation and application, meaningful interpretation which can inform study design and data quality decisions, and existing familiarity within the larger behavioral science community. There is also reasonable consensus regarding the thresholds of the rICC that represent acceptable, and excellent, levels of agreement between two measurements (Harper, 1994).

When we applied the rICC to salivary analyte data, our findings revealed exceptionally high repeatability for all analytes with all rICC estimates exceeding 0.95. At this level of precision, the variability in our salivary data attributed to laboratory-based measurement error is < 5% with the remaining > 95% of variability reflecting differences in analyte levels across samples. Our rICC estimates indicate similarly low levels of measurement error for all our analytes despite variability in average intra-assay CVs across analytes and some instances of high across-replicate variability for individual biospecimens, as indicated by the percentage of samples with intra-assay CVs above 15%. These patterns of findings illustrate one of the added benefits of using the rICC to contextualize the contribution of measurement error relative to variability in the overall analyte data- a comparison that is not possible when measurement error is expressed at the individual biospecimen level (i.e., as with the intra-assay CV). This contextualization of measurement error is especially important when modeling changes and/or differences in salivary biomeasure concentrations associated with hypothesized predictors, group memberships, or independent variables. For example, if researchers find that an intervention effect accounts for 10% of the variance in a salivary analyte's concentration, and the study employed singlet testing with an analyte with a "known" rICC of 0.95, up to half of the "intervention effect" could be due to measurement error. While it is highly unlikely that all the remaining analyte variance in this example (5%) could be attributed to measurement error alone, indexing and contextualizing measurement error in this way provides a more measured understanding of model results. This is especially important when the estimated effects are relatively small. The rICC allows for this contextualization and aids in the interpretation of the overall study findings and the implications of these findings.

The rICC also provides distinct advantages when investigators have identified confounding variables and/or covariates that are important in the calculation of measurement precision (e.g., sampling method or sample quality characteristics) as these variables can be included in the rICC estimation. In contrast, the intra-assay CV, by nature of its measurement at the biospecimen level, is especially valuable when QA/QC procedures are essential during the assay process (e.g., when there is high variability in a study sample or when laboratory staff are inexperienced). Also, when the sample size is very small or the assumptions of an rICC model are not met, intra-assay CV calculations are more appropriate. Assessing both the rICC and individual biospecimen intra-assay CVs in salivary bioscience studies therefore provides a more nuanced understanding of laboratory-based measurement error and its potential impact on the study results and implications.

The high level of precision demonstrated in salivary assay data is rarely required, nor expected, for self-report, interview, or behavioral observation data collected in biobehavioral research. Achieving this high bar for precision has several noteworthy implications. When the rICC indicates that laboratory-based measurement error is minimal, researchers may choose to reduce the number of technical replicates and instead prioritize biological replicates. Testing only a small percentage of technical replicates allows more data to be generated from finite saliva sample volumes and study resources. Shifting resources conventionally allocated for technical replicates (time, funds, disposable supplies, assay reagents) to support assaying more samples per participant, or additional salivary analytes per sample, can enable more complex study designs and research questions. Singlet testing also increases the possibility that sample volume can be reserved for archiving or biorepositories.

Researchers can use the rICC as an objective index to assist in these decision-making processes during the planning, piloting, and implementation phases of their projects. At the earliest stages of study design and proposal, documenting that the salivary assay protocols to be applied yield high rICC indices in a request for funding may add substantial value during the review process. This information would provide the context for an investigative team to objectively justify reducing the number of technical replicates. Reduced replicate testing can make resources available to improve the study design to minimize other, perhaps more important, sources of unsystematic variability (e.g., variability due to sample collection times or collection type). During the planning stages, investigators can also use laboratory- and analyte-specific rICC estimates to select a testing laboratory, refine assay testing protocols, and inform the selection of measurement panels, sampling schemas, and sample size determinations. Once preliminary data are available, rICC estimates can be calculated using these data, and this information can help researchers finalize study plans.

For investigators who have already completed data collection, monitoring or requesting rICC information from technicians conducting their biological testing may also be useful. The rICC indicates how well the assay performs in the hands of the operator. If the data from a particular assay, student, staff member, or laboratory yield a less than exceptional rICC, the investigator may need to adjust testing plans to support a higher percentage of technical replicates, use the rICC as a criterion to rework staff training, or, in the worst-case scenario, select another laboratory with above-threshold performance metrics.

When evaluating rICC estimates calculated from pilot data or reported by testing laboratories or specific operators, the simulation study findings presented here provide valuable information about expected variability in estimated rICCs. For example, if a researcher planning a large-scale salivary biomeasure study estimates a rICC of 0.95 using pilot data from 50 participants, they can be reasonably confident the rICC of their final analyte data will range from 0.92 to 0.98. Depending on the level of precision needed for the specific research question, the investigator may decide to test all or only a proportion of samples in replicate. For laboratories interested in using rICC estimates to help guide assay testing recommendations and demonstrate laboratory measurement precision, our findings suggest that very stable rICC estimates can be calculated using samples of 200 biospecimens, and smaller sample sizes may provide adequate stability in rICC estimates

depending on the specific research question, potential clinical applications, and desired measurement precision. Sample sizes beyond 200 provided only marginal added benefits in rICC measurement precision. Our findings also suggest that laboratories and assay operators should aim to meet our preliminary lower-limit threshold of 0.95 for the rICC, as this criterion was surpassed for all analytes and conditions tested.

Regarding the statistical estimation of the rICC, there are several guidelines we believe should be considered by salivary bioscience researchers. First, we highlight the importance of checking the distribution of the analyte data prior to rICC estimation and evaluating model fit parameters after calculation. While we present rICCs estimated using both raw and log-transformed data, it is important to note that the validity of the rICC estimate could be sensitive to model fit, and non-normal distributions of salivary analyte data may compromise these indices (Schielzeth et al., 2020). Also, although not evaluated in the current study, the estimation of the rICC can be adapted to adjust for additional sources of variation in salivary measurements. For example, rICC models can control for confounding covariates thought to affect measurement precision such as biospecimen collection method and mucin content. In addition, multiplexing technology allows for the testing of multiple analytes in a single well. This introduces another level of variability to the rICC multilevel calculation and allows for shared variance at the level of the plate well. Similarly, future studies could examine adjustments to the rICC calculation that could account for multiple saliva sample collections per participant in the dataset. Modeling the shared variance among biospecimens from the same participant may strengthen rICC estimates. These approaches should be considered when estimating rICCs. Such modifications to the rICC calculations are important areas of investigation for future studies as they may increase the rICC estimates, along with their utility and reliability, for salivary bioscience researchers and testing laboratories.

4.1. Limitations

When applied to salivary biomeasure data, there are several sources of variance that could affect the level and stability of estimated rICCs. In this study, we address variation related to assay type, the distribution of the data, and sample size. However, it is important to note that an analyte's rICC is specific to the testing laboratory, its equipment, and staff. The quality of the laboratory equipment and experience and performance of the laboratory staff are critical factors influencing the precision of analyte determinations. Laboratories and investigators conducting assay work should verify that their equipment is calibrated and in good working order. These factors will directly affect the level of the rICC estimated. Therefore, the rICCs we report should not be generalized beyond our laboratory. We were also not able to assess the sensitivity of rICC estimates to biospecimen quality and characteristics. Various factors related to saliva sample collection, storage, and constitution (e.g., swab vs. passive drool collection, cold chain procedures, and mucosal content) may increase the risk of within-biospecimen variation, thereby reducing the rICC. Future research is needed to understand the effects of these factors on the rICC estimate. Finally, when evaluating the rICC estimates presented in this paper and calculated by other studies or laboratories, it is important to consider how differences in between-biospecimen variance will impact rICC estimates, even when the variance associated with measurement error is held constant (see (Dochtermann and Royauté, 2019) for additional discussion). While we found similarly high rICC

estimates for all our analytes regardless of their level of variation (as can be seen in the wide range of SDs in our analyte data), this is likely due to the small contribution of across-replicate variation in our analyses. When using the rICC to assess biomeasure precision, future researchers should consider both the between-biospecimen and across-replicate variance estimates as well as the rICC estimate.

5. Conclusions

Estimates of the rICC demonstrated the exceptionally precise measurement of salivary analyte concentrations in our example data. The calculation of this additional index of biomeasure precision was easily implemented, and estimates were robust across assay and data characteristics. With additional studies that advance the use and utility of the rICC in salivary bioscience research, the further adoption of this index in biomeasure investigations could support the efficient allocation of study resources and the implementation of more complex and rigorous studies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Bates D, Maechler M, Bolker B, Walker S, Christensen R, Singmann H, Dai B, Scheipl F, Grothendieck G, Green P, Fox J, 2020. lme4: Linear Mixed-Effects Models using "Eigen" and S4.

Blair C, Granger D, Willoughby M, Mills-Koonce R, Cox M, Greenberg M, Kivlighan K, Fortunato C, 2011. Salivary cortisol mediates effects of poverty and parenting on executive functions in early childhood. Child Dev. 82, 1970–1984. 10.1111/j.1467-8624.2011.01643.x. [PubMed: 22026915]

Chard T, 1990. An Introduction to Radioimmunoassay and Related Techniques, 4th ed. Elsevier, Amsterdam, Netherlands.

Dochtermann NA, Royauté R, 2019. The mean matters: going beyond repeatability to interpret behavioural variation. Anim. Behav 153. 10.1016/j.anbehav.2019.05.012.

Granger D, Taylor M, 2020. Salivary Bioscience: Foundations of Interdisciplinary Saliva Research and Applications. Springer International Publishing.

Granger D, Kivlighan K, Fortunato C, Harmon A, Hibel L, Schwartz E, Whembolua G, 2007. Integration of salivary biomarkers into developmental and behaviorally-oriented research: problems and solutions for collecting specimens. Physiol. Behav 92, 583–590. 10.1016/j.physbeh.2007.05.004. [PubMed: 17572453]

Granger S, Gaitonde S, 2020. Biomedical research and related applications: current assay methods and quality requirements in oral fluid diagnostics applications. In: Granger D, Taylor M (Eds.), Salivary

Bioscience: Foundations of Interdisciplinary Saliva Research and Applicationsence Foundations of Interdisciplinary Saliva Research and Applications. Springer Nature, Switzerland, pp. 249–262.

Harper D, 1994. Some comments on the repeatability of measurements. Ringing Migr. 15, 84–90. 10.1080/03078698.1994.9674078.

Laurent H, Stroud L, Brush B, D'Angelo C, Granger D, 2015. Secretory IgA reactivity to social threat in youth: relations with HPA, ANS, and behavior. Psychoneuroendocrinology 59, 81–90. 10.1016/j.psyneuen.2015.04.021. [PubMed: 26036453]

Lessells C, Boag P, 1987. Unrepeatable repeatabilities: a common mistake. Auk 104, 116–121.

Nakagawa S, Schielzeth H, 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biol. Rev 85, 935–956. 10.1111/j.1469-185X.2010.00141.x. [PubMed: 20569253]

R. Core Team, 2019. R: A Language and Environment for Statistical Computing.

Riis J, Out D, Dorn L, Beal S, Denson L, Pabst S, Jaedicke K, Granger D, 2014. Salivary cytokines in healthy adolescent girls: intercorrelations, stability, and associations with serum cytokines, age, and pubertal stage. Dev. Psychobiol 56, 797–811. 10.1002/dev.21149. [PubMed: 23868603]

Riis J, Bryce C, Matin M, Stebbins J, Kornienko O, van Huisstede L, Granger D, 2018. The validity, stability, and utility of measuring uric acid in saliva. Biomark. Med 12. 10.2217/bmm-2017-0336.

Riis J, Chen F, Dent A, Laurent H, Bryce C, 2020. Analytical strategies and tactics in salivary bioscience. In: Taylor M, Granger D (Eds.), Salivary Bioscience: Foundations of Interdisciplinary Saliva Research and Applications. Springer.

Rodriguez C, Granger D, Leerkes E, 2020. Testosterone associations with parents' child abuse risk and at-risk parenting: a multimethod longitudinal examination. Child Maltreat. 10.1177/1077559520930819.

Schielzeth H, Dingemanse N, Nakagawa S, Westneat D, Allegue H, Teplitsky C,   eale D, Dochtermann N, Garamszegi L, Araya-Ajoy Y, 2020. Robustness of linear mixed-effects models to violations of distributional assumptions. Methods Ecol. Evol 11, 1141–1152. 10.1111/2041-210X.13434.

Shrout P, Fleiss J, 1979. Intraclass correlations: uses in assessing rater reliability. Psychol. Bull 86, 420–428. [PubMed: 18839484]

Stoffel M, Nakagawa S, Schielzeth H, 2019. rptR: Repeatability Estimation for Gaussian and Non-Gaussian Data.

Vernon-Feagans L, Cox M, Family Life Project Key Investigators, 2013. The Family Life Project: an epidemiological and developmental study of young children living in poor rural communities. Monogr. Soc. Res. Child Dev 78, 1–150. 10.1111/mono.12046.

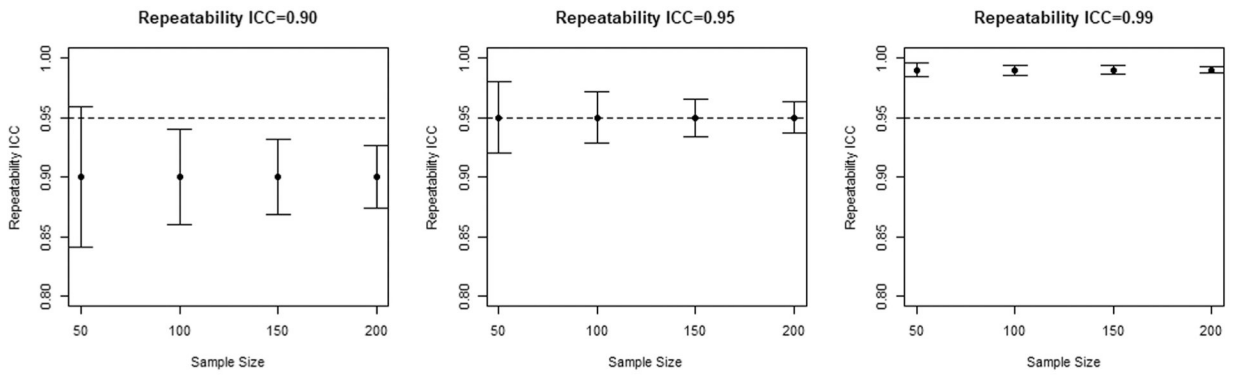**Fig. 1.**
Results from a simulation study assessing the stability of repeatability intra-class correlation (ICC) estimates when calculated using sample sizes from 50 to 200. Note: 1000 bootstrap 95% confidence intervals are presented for each estimated ICC. Dotted lines represent a repeatability ICC of 0.95. See the Methods section for information about the combinations of variance parameters used to generate the simulated data.

Table 1

Resource allocation estimates per salivary analyte tested in duplicate vs. singlet.

Analyte	Tests per Kit		Time to complete assay (h)	Sample testing volume (μL)	
	Duplicate	Singlet		Duplicate	Singlet
Cortisol	38	76	2.0	50	25
Alpha-amylase	46	94	0.5	50	25
C-reactive protein	40	80	3.0	40	20
Interleukin-6	40	80	4.0	50	25
Uric acid	44	88	0.5	20	10
Secretory Immunoglobulin-A	38	76	4.0	50	25
Testosterone	38	76	2.0	50	25

Note: Estimates assume the use of commercially available assay kits.

Table 2

Descriptive statistics for duplicate salivary analyte determinations.

	Assay methodology	Mean (SD)	Median	Range	Skew (kurtosis) of raw data	Skew (kurtosis) of log-transformed data
Cortisol (ug/dL)	ELISA	0.11 (0.10)	0.08	[0.01, 0.78]	3.02 (12.22)	0.11 (0.32)
Alpha-amylase (U/mL)	ELISA	93.13 (66.78)	76.65	[5.25, 368.90]	1.51 (3.00)	−0.82 (0.96)
CRP (pg/mL)	V-Plex	751.75 (3456.78)	137.51	[11.36, 47046.28]	11.50 (145.80)	0.75 (0.66)
IL-6 (pg/mL)	V-Plex	8.31 (13.90)	3.98	[0.45, 136.92]	4.97 (35.30)	0.39 (−0.19)
Uric Acid (mg/dL)	ELISA	2.84 (1.84)	2.44	[0.09, 10.84]	1.32 (2.85)	−1.08 (1.83)
SIgA (μg/mL)	ELISA	177.10 (85.70)	164.13	[15.06, 580.75]	1.21 (2.29)	−0.76 (2.39)
Testosterone (pg/mL)	ELISA	47.91 (30.12)	39.49	[9.56, 255.59]	2.70 (11.27)	0.12 (0.76)

Note: Data from 400 duplicate determinations ($N = 200$) of all analytes except sAA are presented. sAA duplicate determinations were for $N = 60$ (120 duplicate determinations). CRP= c-reactive protein; IL-6 = interleukin-6; SIgA= secretory immunoglobulin-A. See the Methods section for detailed descriptions of assay methodology. Both replicates were used in all calculations.

Table 3

Comparison of intra-assay coefficients of variation (CV) and repeatability intra-class correlations (ICC) for seven salivary analytes tested in duplicate.

	Mean Intra-assay CV% (Median %)	Range of Intra-assay CV	N (%) with intra-assay CV> 5%	N (%) with intra-assay CV> 15%	Repeatability ICC [CI]		
					Using raw data[a]	Using log-transformed data	Using log-transformed and trimmed data[b]
Cortisol	8.00 (5.80)	[0.00, 50.51]	105 (52.50%)	24 (12.00%)	0.990 [0.987, 0.993]	0.976 [0.969, 0.982]	0.982 [0.976, 0.987]
Alpha-amylase	4.97 (2.77)	[0.00, 34.55]	14 (23.33%)	6 (10.00%)	0.985 [0.976, 0.991]	0.990 [0.982, 0.993]	0.997 [0.995, 0.998]
CRP	3.59 (2.57)	[0.00, 78.85]	34 (17.00%)	2 (1.00%)	0.999 [0.999, 0.999]	0.997 [0.996, 0.998]	0.999 [0.999, 0.999]
IL-6	1.91 (1.50)	[0.02, 8.87]	10 (5.50%)	0 (0.00%)	1.00 [0.999, 1.000]	0.999 [0.999, 1.000]	1.00 [0.999, 1.000]
Uric Acid	3.72 (1.59)	[0.00, 69.29]	25 (12.50%)	11 (5.50%)	0.971 [0.962, 0.978]	0.986 [0.981, 0.989]	0.993 [0.991, 0.995]
SIgA	2.52 (1.87)	[0.00, 45.97]	22 (11.00%)	2 (1.00%)	0.997 [0.996, 0.998]	0.992 [0.989, 0.994]	0.996 [0.994, 0.997]
Testosterone	4.50 (3.66)	[0.17, 22.40]	76 (38.00%)	4 (2.00%)	0.987 [0.983, 0.990]	0.988 [0.984, 0.991]	0.989 [0.986, 0.992]

Note: Data from 400 duplicate determinations (N = 200) of all analytes except sAA are presented. sAA duplicate determinations were for N = 60 (120 duplicate determinations). Analyte data are from the first run of assay processing prior to laboratory quality control/quality assurance procedures. CI = 95% bootstrapped confidence interval; CRP= c-reactive protein; IL–6 = interleukin-6; SIgA = secretory immunoglobulin-A.

[a] Raw salivary analyte data were considerably skewed/kurtotic which limits the utility and interpretability of results from models using these non-normal data. This analysis was performed to assess the sensitivity of rICC estimation to the violation of distributional assumptions.

[b] Depending on the analyte, 6–22 potentially influential observations were removed, then rICCs were re-calculated.