The International Journal of Biostatistics

Volume 7, *Issue* 1 2011 *Article* 2

Defining Reproducibility Statistics as a Function of the Spatial Covariance Structures in Biomarker Studies

Irene B. Helenowski, *Northwestern University*
Edward F. Vonesh, *Northwestern University*
Hakan Demirtas, *University of Illinois at Chicago*
Alfred W. Rademaker, *Northwestern University*
Vijayalakshmi Ananthanarayanan, *University of Illinois at Chicago*
Peter H. Gann, *University of Illinois at Chicago*
Borko D. Jovanovic, *Northwestern University*

Recommended Citation:
Helenowski, Irene B.; Vonesh, Edward F.; Demirtas, Hakan; Rademaker, Alfred W.; Ananthanarayanan, Vijayalakshmi; Gann, Peter H.; and Jovanovic, Borko D. (2011) "Defining Reproducibility Statistics as a Function of the Spatial Covariance Structures in Biomarker Studies," *The International Journal of Biostatistics*: Vol. 7: Iss. 1, Article 2.
DOI: 10.2202/1557-4679.1128

Defining Reproducibility Statistics as a Function of the Spatial Covariance Structures in Biomarker Studies

Irene B. Helenowski, Edward F. Vonesh, Hakan Demirtas, Alfred W. Rademaker, Vijayalakshmi Ananthanarayanan, Peter H. Gann, and Borko D. Jovanovic

Abstract

The reproducibility of a biomarker plays a paramount role in determining whether it provides an accurate indication of the true underlying disease or risk status of an individual. When biomarker measurement involves obtaining samples of tissue at random from the organ of interest, sampling variability based on spatial effects can affect this reproducibility. This situation arises when a target organ, such as the prostate or esophagus, is evaluated by multiple random needle biopsies or when an excised organ is randomly sampled. We present a general approach toward estimating reproducibility in the presence of different variance-covariance structures needed to account for possible spatial or temporal variation and correlation. Specifically, we extend the work of previous authors involving applications of the concordance correlation coefficient (CCC) by allowing for different variance-covariance structures of the data. A general concordance correlation matrix representing pairwise concordance correlation coefficients is presented along with an overall concordance correlation coefficient both of which may be obtained from models assuming different variance-covariance structures. The overall concordance correlation coefficient provides a measure of the overall reproducibility and its validity relative to various assumed covariance structures can be assessed by examining commonly employed goodness-of-fit measures. We illustrate these methods to minichromosome maintenance protein 2 (MCM2) data coming from the prostate glands of seven subjects having prostate biopsies between 2002 and 2003.

KEYWORDS: concordance correlation coefficient (CCC), unstructured covariance, compound symmetry, spatial linearity

Author Notes: The authors gratefully acknowledge Ryan J. Deaton, B.S., B.A. (University of Illinois - Chicago), Lawrence I-Kuei Lin, Ph.D. (Baxter Healthcare Corporation), and Tomasz K. Helenowski, M.D. (Stereotactic Radiosurgery Institute) for their contributions used throughout this manuscript. This work was supported by the Prostate SPORE (NIH/NCI P50 CA 90386) and Breast SPORE (P50 CA 89018) grants as well as the NIH/NCI R01 CA 90759 grant through the Robert H. Lurie Comprehensive Cancer Center of Northwestern University.

1. Introduction

In this paper, we present a general framework for evaluating both pairwise and overall reproducibility for a biomarker measured on a spatial continuous scale within the context of structured variance-covariance matrices. We illustrate our approach by examining how distance between two sections of the prostate gland affects the overall reproducibility of the MCM2 biomarker, as determined from measuring their concentrations in the needle core samples taken from each of the sections. Specifically, we assess reproducibility of the MCM2 biomarker under one of five variance-covariance structures, unstructured, compound symmetry, spatial linearity, the spatial spherical structure, and the spatial exponential structure. We determine how appropriate each covariance structure is for evaluating reproducibility by comparing the goodness-of-fit of each model using the likelihood ratio test (Hedeker and Gibbons, 2006). We further conducted simulation studies involving data of different sample sizes generated under different distributions and the mentioned variance-covariance structures to examine the validity of our method under different conditions.

Our methods can be applied in studies involving biopsies of other glands and organs, tissue microarrays, and other methods used in obtaining biomarker data. Adequacy of obtaining biomarker data from biopsies has been discussed by Hewitt *et al*. (2004), where the authors note that renal biopsies provide more information on biomarkers associated with kidney disease than might urine or serum samples. Because of the organ's complexity, this procedure is now only warranted for the most severe cases. If the reproducibility is consistent across the kidney, however, then investigators might perform the procedure on an isolated area to determine biomarker levels in less severe cases as well. Arguments for our methods could also apply to stereotactic needle biopsies done in the brain to study biomarkers. If the biomarker concentrations obtained from the biopsies are similar across different portions of the brain, then fewer biopsies might be required to determine the highest tumor grade, which is used to select treatment modalities for malignant brain tumors (Helenowski, 2006).

Investigators working with ductal lavage fluid in breast cancer research can likewise employ our methods to determine the reproducibility of the fluid across the breast is reproducible. Francescatti *et al*. (2005) and Johnson-Maddux *et al*. (2005) state how ductal lavage fluid might not be obtained from every duct sampled, thus adding to the necessity of determining this reproducibility. Ductal lavage requires much delicacy and precision (2005), leading to another reason why investigators would want to perform this procedure on fewer locations of the breast with little sampling error. Johnson Maddux *et al*. (2005) further discuss

poor reproducibility of the breast fluid over time. We could thus extend our methods to determine how the decrease in reproducibility relates to the time continuum itself and if reproducibility over time will improve with advanced sampling techniques.

2. Statistical Background

In this work, we first discuss the background of already established methods for measuring reproducibility and its dependency on continuous variables and then present how our method adds to these approaches by evaluating how these measures change with respect to different pre-specified variance-covariance structures. When evaluating reproducibility over say K sources (e.g., K observers or raters), pairwise concordance and overall concordance, as initially described by Lin (1989; 2000), are given, respectively, by:

$$\rho_{C_{ij}} = \frac{2\sigma_{ij}}{\sigma_i^2 + \sigma_j^2 + (\mu_i - \mu_j)^2} \tag{1.1}$$

$$\rho_{C_{Overall}} = \frac{2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sigma_{ij}}{(K-1)\sum_{i=1}^{K}\sigma_i^2 + \sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\mu_i - \mu_j)^2} \tag{1.2}$$

for any two samples i and j, where $i = \{1,...,K-1\}$ and $j = i+1 = \{2,...,K\}$ denote the i^{th} and j^{th} of K sources, respectively. These statistics are comprised of the population variance parameters σ_i^2 and σ_j^2, the sample covariance, σ_{ij}, and the sample mean parameters, μ_i and μ_j, of the biomarker samples from sources i and j, respectively.

These measures have been adapted to address various issues related to the assessment of reproducibility. Most notably, Barnhart and Williamson (2001) and Barnhart *et al*. (2001) discuss the applications of generalized estimating equations (GEE) in evaluating the impact of different covariates on pairwise and overall concordance correlation coefficients (CCC's). Barnhart and Williamson (2001) describe a set of three different GEE for obtaining estimates of pairwise CCC's by modeling the means, variances, and CCC's, respectively. They note that an advantage of the GEE approach involves the relaxed assumptions of the distribution of the data and its ability to factor out other sources of inter-subject

DOI: 10.2202/1557-4679.1128

variability accounted for by fixed known sources such as gender or age., a facet of a between-subject problem.

In related work, Carrasco and Jover (2003) use a variance-components approach to examine reproducibility among several measurements by demonstrating the equivalence of the intra-class correlation coefficient (ICC) to the overall CCC (1.2). They show that under a simple linear mixed-effects model with observers treated as fixed effects and subjects as random effects, the ICC will equal the overall CCC of Lin (1989; 2000). If restricted to two observers, the ICC will be equivalent to the pairwise CCC (1.1). Following Barnhart and Williamson, Carrasco and Jover (2003) extend their approach to account for known inter-subject sources of variation by including additional between-subject covariates into their linear mixed-effects model. Unlike the work of these previous authors, we focus on the problem of accounting for spatial within subject variation when estimating pairwise and overall reproducibility – our example examining the spatial effect of distance on the reproducibility between any two prostate gland sections. Our approach combines the marginal modeling strategy adopted by Barnhart and Williamson (2001) with the flexible linear mixed modeling approach of Carrasco and Jover (2003). This is achieved by simply specifying a marginal linear model having a specified mean structure and a specified marginal variance-covariance structure that can accommodate spatial and temporal correlation patterns.

3. Validation Examples

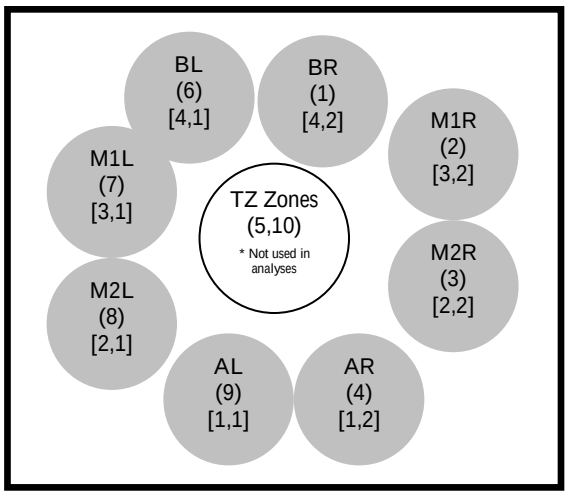
To test the validity of our method for relating within-subject reproducibility to spatial distribution, we applied our model to both real data and simulated data. Applications to real data and accompanying statistical methods are described in Sections 4. Simulation procedures and results are described in Section 7. The implications of these results are discussed in Section 8.

4. Data Example

For our example, we use data for the MCM2 biomarker, coming from seven patients, ages 47 – 68, having prostate biopsies between 2002 and 2003. For the biopsies, needle core samples were taken from each of the ten sections of the prostate. Figure 1 shows the orientation of each section. Our analyses will pertain to the eight peripheral sections of the prostate gland. Four of these peripheral sections are on the left side of the prostate gland and four sections are on the right side. The two central (transitional) sections from the normal compartment and sections with evidence of pre-neoplastic or cancerous features will not be used in the analyses. We will examine the reproducibility of the

MCM2 index, or the percentage of cells taken from each section positively staining for MCM2, between any two given sections of the gland using the pairwise concordance correlation coefficient (CCC) or the overall CCC. Finally, we also will look at the fit of the models from which we derive the elements of the variance-covariance matrix used in the overall CCC to determine the validity of the overall CCC based on the assumed variance-covariance structure**.** All analyses will be applied to data rank-ordered by core ID; normality tests and plots of the data indicated that this approach satisfied residual normality assumptions (Conover and Iman, 1981; Wallis, 1939; Helenowski *et al*., 2003). This approach is equivalent to calculating the Spearman correlation across the subjects between cores, except that we now account for repeated measures. Other transformations, such as log transformation of the data were also tried, but rank-ordered data proved most satisfactory in assuming normality.

Figure 1: Orientation of prostate gland sections from which the data in the example was taken. Note that we use only the data coming from the eight peripheral regions, as opposed to the central or transitional (TZ) sections. The numbers in the parentheses indicate the number of the biopsy needle core designated to take a tissue sample from that particular prostate gland section in each subject. The pair of numbers in brackets represent the spatial coordinates when the prostate is viewed as a 4 x 2 grid consisting of 4 rows and 2 columns. The columns represent the left and right side of the gland. We arbitrarily chose core 9 to represent row 1, column 1 (i.e., coordinate [1,1]). Based on this orientation, core 4 represents row 1, column 2 (i.e., coordinate [1,2]); core 8 represents row 2, column 1 (i.e., coordinate [2,1]), etc.



(1) BR = Basal Right section
(2) M1R = Middle 1 Right section
(3) M2R = Middle 2 Right section
(4) AR = Apical Right section
(6) BL = Basal Left section
(7) M1L = Middle 1 Left section
(8) M2L = Middle 2 Left section
(9) AL = Apical Left section

DOI: 10.2202/1557-4679.1128

5. Statistical Methods

The pairwise and overall CCC given in (1.1) and (1.2) may be estimated directly on the basis of the sample moments from a given pair of readings as shown by Lin (1989). Alternatively, pairwise CCC's and an overall CCC can be computed based on a marginal linear model having a specified mean and variance-covariance structure. Consider the marginal model:

$$y_{km} = \mu + \alpha_k + \varepsilon_{km}, \qquad (5.1)$$

where y_{km} is a biomarker measurement taken from the source effect α_k for $k = \{1,\ldots,K\}$ sources considered as fixed effects (e.g., K different raters or instruments or, in our example, different gland sections), and the ε_{km} are random error terms for $m = \{1,\ldots,M\}$ subjects. The vector of errors for the mth subject, $\boldsymbol{\varepsilon}_m = (\varepsilon_{1m}, \varepsilon_{2m}, \ldots, \varepsilon_{Km})^T$ is assumed to be normally distributed with mean vector $\mathbf{0}$ and variance-covariance matrix, $\Sigma(\theta)$, where θ represents a vector of parameters corresponding to some specified variance-covariance structure. From this model, each subject's vector of readings, $\boldsymbol{y}_m = (y_{1m}, y_{2m}, \ldots, y_{Km})^T$ follows a multivariate normal distribution with mean vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)^T$ where $\mu_k = \mu + \alpha_k$, and variance-covariance, $\Sigma(\theta) = ((\sigma_{kl}(\theta)))_{k,l=1,\ldots,K}$.

Parameters of this model can be estimated for various specifications of $\Sigma(\theta)$, including those with different spatial patterns, using either maximum likelihood (ML) or restricted maximum likelihood (REML) estimation. A concordance correlation matrix, $\mathbf{R}_c$, representing pairwise CCC's between the K different sources can then be calculated directly from the estimated values of $\boldsymbol{\mu}$ and $\Sigma(\theta)$ using (1.1). Likewise, we can compute an overall CCC (OCCC) using (1.2) but corrected for mean bias as indicated by Carrasco and Jover (2003). A macro program to implement this method is available upon request.

By specifying $\Sigma(\theta) = \sigma^2(1-\rho)\mathbf{I} + \rho\mathbf{J}$ where $\theta = (\sigma^2, \rho)$, $\mathbf{I}$ is the $K \times K$ identity matrix and $\mathbf{J}$ is the $K \times K$ unit matrix of 1's, model (3.1) is the marginal form of the linear mixed-effects model of Carrasco and Jover (2003) with ρ representing the usual ICC obtained under a simple random intercept mixed-effects model. In this case, the concordance correlation matrix, $\mathbf{R}_c$, will reflect the assumed compound symmetric covariance structure and any differences between the pairwise CCC's will reflect differences in location shift between the various sources, α_k. When $\Sigma(\theta)$ is unstructured (i.e., arbitrary positive-definite), the concordance correlation matrix, $\mathbf{R}_c$, will correspond to the $K(K-1)/2$ unique pairwise CCC's that can be obtained using (1.1). For balanced data, these will coincide exactly with the moment-based estimator of Lin (1989) with ML estimation, while use of REML estimation will result in the unbiased moment-

based estimator of Carrasco and Jover (2003). With balanced data, it is straightforward to show that the OCCC obtained under model (3.1) assuming compound symmetry is the same as that OCCC obtained under model (3.1) assuming an unstructured variance-covariance matrix (see appendix for proof). Conceptually, assuming compound symmetry is equivalent to using a weighted average of all the variances with each variance coming from each level of the source effect (e.g., from each biopsy needle core sample in our example). This average is obtained not only across all the source effects but also across all subjects. The common covariance between any two source effects (e.g., needle core samples corresponding to two prostate gland zones within a subject) is likewise obtained via a weighted average across all pairwise combinations of source effects across all subjects. Consequently, since the weights are the same for balanced data, the OCCC in (1.2) may be computed using either an unstructured covariance matrix or a compound symmetric covariance matrix. This approach has important implications in those cases where one has highly unbalanced or sparse data, for example. For a data set involving K = 8 and M = 7, an 8x8 unstructured variance-covariance matrix may not be estimable, for instance. We therefore rely on the compound symmetric structure to draw inference on the OCCC. One advantage of using model (3.1) to obtain an overall reproducibility statistic is that fitting the data assuming a covariance structure that may be biologically meaningful, as in the case of applying model (3.1) assuming spatial linearity, for instance. In this manner, we can examine both pairwise and overall reproducibility of the biomarker taking into account spatial distance. A model that incorporates a spatial pattern assumes that the covariance between biomarker samples from any two sources depends on this continuous variable whether the dependency exists or not. Spatial covariance structures can incorporate dependency on distance via an exponential spatial covariance structure, a spatial power structure, or a spherical structure. For our analyses, we try spatial covariance structures, where the covariance will depend linearly, spherically, or exponentially on the Euclidean distance between two prostate sections, based on a rectangular spatial coordinate system, as shown in Figure 1. The coordinates correspond to a 4×2 grid consisting of 4 rows and 2 columns. The columns represent the left and right side of the gland. We arbitrarily chose core 9 to represent row 1, column 1 (i.e., coordinate [1,1]). Based on this orientation, core 4 represents row 1, column 2 (i.e., coordinate [1,2]); core 8 represents row 2, column 1 (i.e., coordinate [2,1]), etc. For example, the Euclidean distance between core 1 and core 9 is $\sqrt{(2-1)^2+(4-1)^2} = \sqrt{10}$ = 3.16228. Euclidian distance is used in this work, since we are interested in relationships between prostate gland zones corresponding to needle cores only with respect to their relative coordinate positions and not to their absolute locations. Previous work has mentioned how this Euclidean distance approach serves as an equivalent

alternative to functions like the semi-variogram or variogram in spatial statistics (Atkinson and Lewis, 2000; Wagner, 2003). The specific form for the spatial linear covariance matrix for sources k and l, for example, is

$$Cov(Y_{km}, Y_{lm}) = \sigma^2\{1-(\rho\, d_{kl})\}\times I\{\rho\, d_{kl} \le 1\} \qquad (5.2)$$

where d_{kl} is the Euclidean distance between cores k and l based on the spatial coordinates of Figure 1, and $I(\rho\, d_{kl} \le 1)$ is the indicator function taking a value of 1 when $\rho\, d_{kl} \le 1$, and 0 otherwise. If pairwise reproducibility does depend on distance via the assumed structure in (3.2), then the pairwise reproducibility statistics (or CCC's) will decrease with increasing distance by a factor of $1 - \rho$. Consequently, the reproducibility between samples coming from two sources decreases with increasing Euclidean distance between the sources.

For the data discussed in this work, we will consider the five variance-covariance structures of unstructured covariance (UN), compound symmetry (CS), spatial linearity (SL), the spatial spherical structure (SS), and the spatial exponential structure (SE). Since we can not fit an unstructured variance-covariance matrix to all 7 subjects due to sparseness of data ($K>M$), we obtained pairwise CCC's by fitting model (3.1) to each pair of sources and from these separate model fits estimated the CCC for that pair. We then constructed a concordance correlation matrix based on these individual paired fits. We also compute both pairwise and overall CCC's using model (3.1) assuming the variance-covariance structures of compound symmetry, spatial linear, spatial spherical, and spatial exponential structures, however.

To determine which covariance structure best describes the true relationship between overall reproducibility and distance, we used the likelihood ratio test (LRT), involving a χ^2 distribution with degrees of freedom determined by the number of covariance parameters, as described in Hedeker and Gibbons (2006). We also examined graphically, the appropriateness of each covariance structure by plotting the pairwise CCC's from (1.1) versus distance based on our coordinate system, superimposing the average pairwise CCC's for each covariance structure assumption. Such checks on goodness-of-fit of an assumed covariance structure are paramount as different covariance structures can provide similar estimates of overall reproducibility (OCCC) but still have markedly different estimates of pairwise reproducibility. Inference regarding the OCCC can be carried out using methods described by Carrasco and Jover (2003) but applied to our model (3.1), where standard error estimates can be obtained via the delta method.

6. Results

Table 1 gives the means and standard deviations for the untransformed MCM2 data. Table 2 contains the individual pairwise CCC's for each unique pair of cores obtained assuming unstructured covariance, compound symmetry, or one of the spatial structures mentioned above applied to the data. As described above, the pairwise CCC's for an unstructured covariance matrix were obtained strictly on a pairwise basis as we could not fit model (3.1) assuming an overall unstructured covariance across all 8 cores. This unstructured concordance correlation matrix is presented strictly for comparative purposes. For the MCM2 data, the pairwise CCC's obtained assuming unstructured covariance and averaged within Euclidean distance for this data set range from 0.00 to 0.57, with 67% of the correlations being greater than 0.50. Again noting the equivalence of these CCC's to the ICC, we consider the majority of these pairwise reproducibility measures as fair to good. Rosner (2006) states that a value around 0.40 for the ICC, and thus for the equivalent form of the CCC, indicates fair to good reproducibility.

This is further evidenced in Figure 2 where we plot the individual unstructured pairwise CCC's from Table 2 (corresponding to an unstructured covariance matrix) versus Euclidean distance. Superimposed with connecting lines are the average pairwise CCC's from Table 2 at each Euclidean distance. As illustrated, there is no evidence of a spatial trend; thus compound symmetry appears to provide a reasonable fit to the pairwise unstructured CCC's.

Table 1: Core means and standard deviations for MCM2 data.

Real Data			
CoreID	N	Mean	Standard Deviation
1	7	35.74	13.32
2	7	33.99	6.62
3	7	36.36	7.77
4	6	36.90	12.68
6	6	34.43	6.61
7	7	35.25	7.22
8	7	34.83	7.34
9	7	34.39	4.06

DOI: 10.2202/1557-4679.1128

Table 2: Concordance correlation matrices based on unstructured covariances, compound symmetry, spatial linearity, the spatial spherical structure, and the spatial exponential structure for MCM2 data. The concordance correlation coefficients for the unstructured covariance case were obtained by pairwise fitting of model (3.1) to each pair of cores (or equivalently, by directly estimating from equation (1.1) using the sample moments).

		Core							
			1	2	3	4	6	7	8	9
Unstructured Covariance under model (3.1)	*Core*	1	1.000	0.068	0.530	0.897	0.522	0.421	0.532	0.211
		2		1.000	0.286	0.048	0.000	0.154	0.650	0.584
		3			1.000	0.726	0.560	0.327	0.663	0.568
		4				1.000	0.642	0.340	0.659	0.260
		6					1.000	0.758	0.689	0.480
		7						1.000	0.398	0.527
		8							1.000	0.555
		9								1.000
Compound Symmetry under model (3.1)	*Core*	1	1.000	0.570	0.570	0.559	0.533	0.570	0.570	0.570
		2		1.000	0.570	0.559	0.533	0.570	0.570	0.570
		3			1.000	0.559	0.533	0.570	0.570	0.570
		4				1.000	0.562	0.559	0.559	0.559
		6					1.000	0.533	0.533	0.533
		7						1.000	0.570	0.570
		8							1.000	0.570
		9								1.000

Table 2, continued

		Core							
		1	2	3	4	6	7	8	9
Spatial Linearity under model (3.1)	*Core*	1	1.000	0.708	0.417	0.124	0.676	0.587	0.348	0.078
		2		1.000	0.708	0.415	0.561	0.708	0.587	0.348
		3			1.000	0.705	0.332	0.587	0.708	0.587
		4				1.000	0.076	0.346	0.585	0.705
		6					1.000	0.676	0.398	0.119
		7						1.000	0.708	0.417
		8							1.000	0.708
		9								1.000
Spatial Spherical under model (3.1)	*Core*	1	1.000	0.602	0.265	0.048	0.572	0.452	0.201	0.028
		2		1.000	0.602	0.263	0.429	0.602	0.452	0.201
		3			1.000	0.597	0.190	0.452	0.602	0.452
		4				1.000	0.028	0.199	0.448	0.597
		6					1.000	0.572	0.251	0.046
		7						1.000	0.602	0.265
		8							1.000	0.602
		9								1.000

DOI: 10.2202/1557-4679.1128

Table 2, continued

Spatial Exponential under model (3.1)	*Core*	*Core*							
			1	2	3	4	6	7	8	9
		1	1.000	0.566	0.320	0.179	0.530	0.447	0.280	0.165
		2		1.000	0.566	0.316	0.418	0.566	0.447	0.280
		3			1.000	0.559	0.262	0.447	0.566	0.447
		4				1.000	0.161	0.276	0.442	0.559
		6					1.000	0.530	0.300	0.169
		7						1.000	0.566	0.320
		8							1.000	0.566
		9								1.000

The overall CCC's in Table 3 give values among all sections of the prostate gland for the MCM2 data assuming either compound symmetry (0.61), spatial linearity (0.54), the spatial spherical structure (0.42), and the spatial exponential structure (0.44). From the -2 log likelihood values in Table 3, we see that the model assuming compound symmetry provides the best fit for the data. The LRTs indicate that this model provides a significantly better fit than the models assuming spatial linearity ($p = 0.004$), a spatial spherical structure ($p = 0.01$), or a spatial exponential structure (0.055).

Table 3: Overall Reproducibility (Overall CCC) and -2 log likelihood values and likelihood ratio tests (LRTs) for models assuming different variance-covariance structures when applied to the MCM2 data.

Model Assuming	**Overall CCC**	**-2log Likelihood**	**p-value vs. Compound Symmetry**	**p-value vs. Spatial Linearity**	**p-value vs. Spatial Spherical**
Compound Symmetry	0.61	190	-	-	-
Spatial Linearity	0.54	198	0.0041	-	-
Spatial Spherical	0.42	196	0.0099	0.2092	-
Spatial Exponential	0.44	193	0.0547	0.0333	0.0857

7. Simulation Approach and Summary of Results

To examine the effectiveness of our method on other data sets of various sizes, and following different distributions we also simulated data assuming a nested design with observations nested within "subjects". Parameters used in generating normal data included 0, 15, and 30 for the mean, and 1, 2, 5, and 10 for the variance. In generating gamma distributed data, 15 and 30 were considered for the location parameter and, and 1, 2, 5, and 10 were considered for the scale parameter. Finally, data following a Uniform(0,1) distribution were also generated. Different samples sizes were likewise considered in our simulation study involving 5, 8, or 10 observations per subject for 5, 10, 25, or 50 subjects. Pre-specified variance-covariance structures used in generating the data included

DOI: 10.2202/1557-4679.1128

compound symmetry, spatial linearity, the spatial spherical structure, and the spatial exponential structure; these structures were associated with initial correlations of 0.1, 0.5, or 0.9.

Under each distribution and variance-covariance structure, we randomly selected 30 parametric settings with which we generated 1000 data sets and applied our model. Only random selections of parametric settings were tested given each distribution and pre-specified variance-covariance structure, as all possible combinations of location, scale, and correlation parameters could not be feasibly tried.

Results are presented in Table 4. For each combination of data distribution and prespecified covariance structure, the intent was to run 1000 simulations for each of the randomly selected parametric settings. The third column of Table 4 gives the numbers and percentage (out of 30) of settings where all 1000 simulations converged. For each simulation, the model was fit using four covariance assumptions. The fourth column of Table 4 gives the percentage (out of the total number of simulations) for which the model with the highest likelihood had a covariance structure that matched with the prespecified covariance structure.

In general, the probability of agreement was slightly lower for data generated under a spatial variance-covariance structure as opposed to under a compound symmetric variance-covariance structure, however. The probability of agreement also tended to be smaller for data with scalar parameters (variance) being relatively smaller than the location parameters (mean) and with smaller pre-specified correlation parameters in normal data. Similar results were also observed for data following the gamma distribution. Contrarily, the probability of agreement tended to be greater with an increasing magnitude of the pre-set correlation parameter in uniform data as well as in normally distributed and gamma distributed data. We note here that results were similar for generated data sets of different sample sizes tried.

Table 4: Summary of simulation results indicating the distribution that the data was generated under, the pre-specified variance-covariance structure used in generating the data, the percentage of data settings (out of 30) for which models under all four variance-covariance structure considered converged for all 1000 simulations, the percentage of convergent simulations where the variance-covariance structure (VC) indicated by our method as providing the optimal fit for the data was the same as the pre-specified VC, and the total number of convergent models.

Distribution of Data	Pre-specified Variance-Covariance Structure (VC)	Number and Percentage (%) out of 30 Convergent Models	Percentage (%) of Convergent Models (out of the Total Number of Convergent Simulations) in agreement with the Pre-specified VC's	Total Number of Convergent Simulations
Normal	Compound Symmetric	28 (93.33%)	96.43%	28,000
	Spatial Linear	20 (66.67%)	93.10%	20,000
	Spatial Spherical	18 (60.00%)	94.50%	18,000
	Spatial Exponential	16 (53.33%)	93.90%	16,000
Gamma	Compound Symmetric	28 (93.33%)	96.43%	28,000
	Spatial Linear	20 (66.67%)	94.40%	20,000
	Spatial Spherical	19 (63.33%)	95.00%	19,000
	Spatial Exponential	19 (63.33%)	95.20%	19,000
Uniform	Compound Symmetric	27 (90.00%)	94.60%	27,000
	Spatial Linear	20 (66.67%)	94.80%	20,000
	Spatial Spherical	21 (70.00%)	100.00%	21,000
	Spatial Exponential	28 (93.33%)	97.50%	28,000

DOI: 10.2202/1557-4679.1128

8. Conclusion

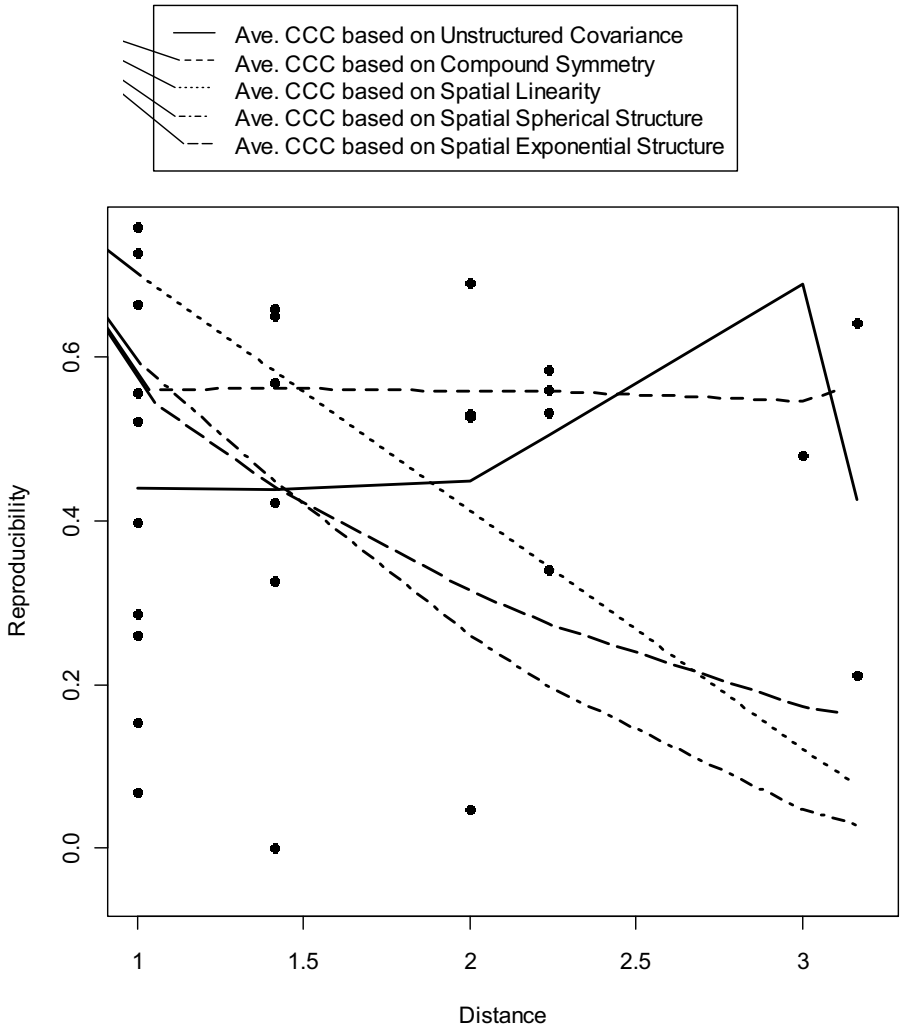
The statistical methods implemented here can help investigators in determining how a continuous variable as spatial distance affects a measure of biomarker reproducibility. Such procedures may prove useful to medical researchers examining biomarkers thought to be associated with the diseases they are studying. Carrasco and Jover (2003) first discussed the equivalence of the concordance correlation coefficient (CCC) given in Lin (1989; 2000) to the intraclass correlation coefficient (ICC), applying elements of a variance-covariance structure from a mixed-effects model to these statistics. As a result, they were able to describe reproducibility using variation of fixed effects as well as random effects. We use a strictly marginal model allowing for different variance-covariance structures that includes the approach in Carrasco and Jover (2003), where the marginal model is obtained by integrating out the random effects. Our approach is slightly more flexible in that it allows us to examine the impact of various assumed covariance structures and not simply a random-effects structure. Moreover, it allows one to investigate spatial and temporal variation effects on reproducibility.

Applying our methods to MCM2 data from peripheral regions of the prostate gland, we determined that the overall CCC values obtained from the model assuming compound symmetry (0.61) was higher than the overall CCC values obtained from the models assuming any of the spatial structures (0.42 to 0.54), as shown in Table 3. These results indicate that the overall reproducibility among the cores is independent of the distance between them, since once again the model assuming compound symmetry offers the best fit for the data.

Models obtained assuming the spatial structures indicate a poorer fit to the data, as indicated by the likelihood ratio test (Table 3) and by visual inspection of the MCM2 data (Figure 2). The lesser fit using spatial dependency indicates that there is consistent reproducibility across sections, and that this reproducibility is better represented by the compound symmetry model.

Simulation results for the normal, gamma, and uniform data of different samples sizes also has shown that our method indicates that the variance-covariance structure that the data is generated under as providing the optimal fit of the data in most cases. Therefore, our method generally leads to correct inferences concerning spatial dependence among within-subject observations. This inference, however, can be influenced by the relationship between location and scale parameters associated with the distribution of the data and magnitude of the pre-set correlation parameter. These occurrences may results from variation being concentrated on between-subject differences, as within-subject variation approaches zero.

Figure 2: Superimposing the average pairwise reproducibility measures (CCC's) versus Euclidean distance for the original data obtained from the models assuming unstructured covariance, compound symmetry, spatial linearity, the spatial spherical structure, and the spatial exponential structure. With assuming the spatial structures, we see a downward linear trend across distance, corresponding to a coefficient $\rho > 0$ as presented in equation (3.2). Individual pairwise CCC's (indicated by the points) indicate no clear relationship between reproducibility and distance, consistent with the results coming from the models assuming unstructured covariance or compound symmetry.



DOI: 10.2202/1557-4679.1128

Barnhart *et al*. (2001) suggest a scheme incorporating different weights for the pairwise CCC's when calculating of the overall CCC's, which could have affected the overall CCC calculations based on the model assuming a variance-covariance structure dependent on distance. The aim of our study was created to test uniform biomarker concentrations throughout the prostate gland, however. Therefore, calculations of the overall CCC using equal weights are appropriate. As mentioned in the introduction, these methods can also be extended to several studies involving similar designs such as biopsies of other organs, tissue microarrays, and ductal lavage procedures. These studies demonstrate a potential need for methods to determine the dependence of biomarker reproducibility on a continuous variable as distance and could be extended to a continuous variable such as time also.

The approach we take can be generalized to include between-subject covariates as in Barnhart and Williamson (2001) and Barnhart *et al*. (2001). These covariates may be included directly in model (3.1) in the same way Carrasco and Jover (2003) do in their linear mixed-effects model. One advantage of the approach taken by Barnhart and Williamson (2001) is that they allow modeling of the CCC directly as a function of covariates. Nevertheless, the approach of Barnhart *et al*. (2001) requires specifying a complex set of estimating equations each making use of the normality assumption in order to specify the higher order working covariance matrices needed to fit each set of GEE. Vonesh *et al.* (2001) overcome these issues by assuming normality and using a marginal linear model with specified covariance structure in combination with robust standard errors to obtain the same level of flexibility described by Barnhart and Williamson (2001). Specifically, Vonesh *et al*. (2001) show that using the maximum likelihood estimation approach under the normality assumption is equivalent to solving a set of second-order GEE (GEE2) under non-Gaussian assumptions. Using an empirical sandwich estimator can provide robust standard errors of the CCC by applying the delta-method in a manner similar to that in Carrasco and Jover (2003).

The coverage probability (CP) and total deviation index (TDI) measures mentioned in Lin *et al*. (2006), along with the CCC, can also be used to measure reproducibility between individuals for each prostate gland zone in our example. These additional indexes, TDI (δ) and CP (π) allow us to obtain information for a large portion of the data, without information on total data variability. Consequently, these two indexes can be explored in future work regarding the dependence of reproducibility on a continuous variable given limited data. Lin *et al*. (2006) provided simulation studies which showed that given k subjects, if m, the number of measures increases to infinity, these proposed indexes match those proposed by Barnhart *et al*. (2001). If m = 1, on the other hand, the indexes match with the ICC values given by Carrasco and Jover (2003). Thus, the work of Lin *et al*. (2006) provides common ground in evaluating reproducibility using different

indexes. We should be aware that different interpretations of the indexes depend on the nature of the data, however. It may also be useful to examine how precision and accuracy are affected by spatial effects.

This paper has presented a number of alternatives to assess reproducibility of markers taken in a spatial pattern. For the MCM2 data analyzed in this paper, reproducibility over the space was indicated by a compound symmetry model being the best fit. Simulation studies in most cases also indicated the variance-covariance structure that data were generated under as providing the optimal fit, further reinforcing the potential effectiveness of our method. By fitting a space-dependant covariance structure such as spatial linearity, the spatial spherical structure, or the spatial exponential structure, other data sets may indicate that markers taken at a greater distance from each other would have less similar levels than those taken closer together, as can be implied from our simulation studies.

Here, we conclude that our method could be applied to any data involving within-subject observations in order to determine which variance-covariance structure provides the optimal fit for those within-subject values. The type of variance-covariance structure determined as providing the best fit for the data may then aid in inference about the type of impact that the spatial distance between within-subject observations have on their reproducibility. Therefore, we recommend this method to investigators working with such data involving this design, attempting to answer such questions.

9. Appendix

The following proofs show the equivalence of the pairwise and overall concordance correlation coefficients (CCC's) derived from variance components obtained via models assuming either unstructured covariance or compound symmetry for balanced data. (A1) shows how the pairwise CCC's are equivalent for the balanced case, while (A2) to (A5) show how the overall CCC's are equivalent for balanced data.

For any given unique pair of sources i and j,

$$\sigma^2 = \frac{\sigma_i^2 + \sigma_j^2}{2}; i = 1,..., K-1; j = 2,...K; i < j$$

which implies

$$\sigma^2 + \sigma^2 = 2\sigma^2 = 2\left[\frac{\sigma_i^2 + \sigma_j^2}{2}\right] = \sigma_i^2 + \sigma_j^2; i = 1,..., K-1; j = 2,...K; i < j \qquad \text{(A1)}.$$

The equation in (A1) shows that the denominator of the pairwise CCC will be the same for balanced data, but not for unbalanced data, regardless of which of these two variance-covariance structures we use.

DOI: 10.2202/1557-4679.1128

We can likewise show the equivalence of the overall CCC's if we were to employ the sample moments obtained using model (1.2). The numerators are the same for balanced data assuming either variance-covariance structure, as shown in (A2) below. Specifically,

$$\sigma = \frac{\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sigma_{ij}}{\left(\frac{K(K-1)}{2}\right)}; i = 1,...,K-1; j = 2,...K; i < j$$

which implies

$$2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sigma = 2\left(\frac{K(K-1)}{2}\right)\sigma = 2\left(\frac{K(K-1)}{2}\right)\frac{\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sigma_{ij}}{\left(\frac{K(K-1)}{2}\right)} = 2\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}\sigma_{ij}; \qquad \text{(A2)}$$

$$i = 1,...,K-1; j = 2,...K; i < j$$

The equivalence of the denominators is given in (A3), (A4), and (A5).

$$\sum_{i=1}^{K}\sigma^2 = K\left(\sigma^2\right) = K\left(\frac{\sum_{i=1}^{K}\sigma_i^2}{K}\right) = \sum_{i=1}^{K}\sigma_i^2; \quad i = 1,...,K, \qquad \text{(A3)}$$

$$\frac{1}{n}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\sigma^2+\sigma^2+2\sigma) = \frac{1}{n}\left(\sum_{j=i+1}^{K}(K-1)\sigma_j^2 + \sum_{i=1}^{K-1}(K-1)\sigma_i^2 + 2K(K-1)\sigma\right)$$

$$= \frac{1}{n}\left((K-1)\sum_{j=i+1}^{K}\sum_{i=1}^{K-1}\frac{\sigma_i^2}{K-1} + (K-1)\sum_{j=i+1}^{K}\sum_{i=1}^{K-1}\frac{\sigma_j^2}{K-1} + 2K(K-1)\sum_{j=i+1}^{K}\sum_{i=1}^{K-1}\frac{\sigma_{ij}}{K(K-1)}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\sigma_i^2+\sigma_j^2+2\sigma_{ij})$$

(A4)

which implies

$$\begin{aligned}&(K-1)\sum_{i=1}^{K}\sigma^2+\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\mu_i-\mu_j)^2-\frac{1}{n}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\sigma^2+\sigma^2+2\sigma)\\&=(K-1)\sum_{i=1}^{K}\sigma_i^2+\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\mu_i-\mu_j)^2-\frac{1}{n}\sum_{i=1}^{K-1}\sum_{j=i+1}^{K}(\sigma_i^2+\sigma_j^2+2\sigma_{ij})\end{aligned} \quad \text{(A5)}$$

10. References

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Trans. Autom. Contr. 19*, 716-723.

Atkinson, P.M. and Lewis, P. (2000) Geostatistical Classification for Remote Sensing: an Introduction. *Computers & Geosciences*. *26*, 361-371.

Barnhart, H.X., Haber, M. and Song, J. (2002). Overall Concordance Correlation Coefficient for Evaluating Agreement Among Multiple Observers. *Biometrics. 58*, 1020-1027.

Barnhart, H.X. AND Williamson, J.M. (2001). Modeling Concordance Correlation via GEE to Evaluate Reproducibility. *Biometrics. 57*, 931-940.

Carrasco, J.L. and Jover, L. (2003). Estimating the Generalized Concordance Correlation Coefficient through Variance Components. *Biometrics 59*, 849-858.

Conover, W.J. and Iman, R. L. (1981) Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician*. *35*, 124 129.

Fabian, C.J., Kimler, B.F., Mayo, M.S. and Khan, S.A. (2005). Breast-tissue sampling for risk assessment and prevention. *Endocr. Relat. Cancer. 12*, 185-213.

Francescatti, D.S., Kluskens, L. and Shah, L. Ductal lavage in the high_risk patient. (2005). *Am. J. of Surgery*. *189*, 340-341.

Hedeker, D. and Gibbons, R.D. (2006) *Longitudinal Data Analysis*. John Wiley & Sons, Hoboken, NJ, pp. 122-129.

DOI: 10.2202/1557-4679.1128

Helenowski, I.B., Jovanovic, B.D., Chatterton, R.T., Geiger, A.S. and Gann, P.H. (2003). Comparison of Parametric and Non-Parametric Methods for Examining the Reproducibility of Breast Fluid Biomarkers. *Proceedings of the American Statistical Association, Biometrics Section*, 1564-1570.

Helenowski, T.K. (16 October 2005). Personal Communication.

Hewitt, S.M., Dear, J. and Star, R.A.,. (2004). Discovery of Protein Biomarkers for Renal Diseases. *J. Am. Soc. Nephrol. 15*, 1677-1687.

Johnson-Maddux, A., Ashfaq, R., Cler, L., Naftalis, E., Leitch, A.M., Hoover, S. and Euhus, D.M. (2005). Reproducibility of cytologic atypia in repeat nipple duct lavage. *Cancer 103*, 1129-1136.

Lin, L. I-K. (1989). A concordance correlation coefficient to evaluate reproducibility. *Biometrics, 45* 255-268.

Lin, L. I-K. (2000). Correction to "A concordance correlation coefficient to evaluate reproducibility". *Biometrics 56*, 324-325.

Lin, L., Hedayat, AS, Wu, W. (2006). A Unified Approach for Assessing Agreement for Continuous and Categorical Data. *J. Biopharm Stat. 17*, 629-657.

Rosner, B. (2006). *Fundamentals of Biostatistics: 6th ed*. Duxbury Press, Pacific Grove, CA, pp. 562 566.

Shen, R. (17 October 2005). Personal Communication.

Vonesh, E.F., Wang, H. and Majumdar, D. (2001). Generalized least squares, Taylor series linearization, and Fisher's scoring in multivariate nonlinear regression. *JASA 96*, 282-291.

Wagner, H.H. (2003). Spatial Covariance in Plant Communities: Integrating Ordination, Geostatistics, and Variance Testing. *Ecology 84,* 1045-1057.

Wallis, W.A. (1939). The Correlation Ratio for Ranked Data. *JASA*, *34*, 533-538.