APPLYING A SPATIOTEMPORAL MODEL FOR LONGITUDINAL CARDIAC IMAGING DATA

BY BRANDON GEORGE[*,1], THOMAS DENNEY, JR.[†], HIMANSHU GUPTA[*,†],
LOUIS DELL'ITALIA[*,‡] AND INMACULADA ABAN[*]

*University of Alabama at Birmingham[*], Auburn University[†] and
Birmingham Veteran Affairs Medical Center[‡]*

Longitudinal imaging studies have both spatial and temporal correlation among the multiple outcome measurements from a subject. Statistical methods of analysis must properly account for this autocorrelation. In this work we discuss how a linear model with a separable parametric correlation structure could be used to analyze data from such a study. The goal of this paper is to provide an easily understood description of how such a model works and discuss how it can be applied to real data. Model assumptions are discussed and the process of selecting a working correlation structure is thoroughly discussed. The steps necessitating collaboration between statistical and scientific investigators have been highlighted, as have considerations for missing data or uneven follow-up.

The results from a completed longitudinal cardiac imaging study were considered for illustration purposes. The data comes from a clinical trial for medical therapy for patients with mitral regurgitation, with repeated measurements taken at sixteen locations from the left ventricle to measure disease progression. The spatiotemporal correlation model was compared to previously used summary measures to demonstrate improved power as well as increased flexibility in the use of time- and space-varying predictors.

1. Introduction. Imaging studies have grown in popularity in recent years as clinical investigators are making use of the ability of imaging modalities to accurately measure outcomes within the body. Traditionally, data collection was typically limited to general outcomes (such as stroke, death, hospitalization) or values observable externally through a physical exam, patient history or blood work. Such outcomes are of interest to both patient and physician, but may be far down the causal pathway from the underlying pathology of a disease. Imaging modalities allow quantification of internal anatomical or physiological properties that would have previously required invasive surgery or autopsy, which provides new understanding of how the body works. There is a hope in the medical community that measures closer to the true pathology of the disease may be more

sensitive to changes thereof, potentially granting additional statistical power for studies [Standaert (2014)].

Considering that, it makes perfect sense for imaging studies to be joined with longitudinal studies, which allow for observation of how an outcome changes over time. Although these longitudinal imaging studies offer wonderful knowledge of how the inside of the body changes over time, there are challenges in the statistical analysis of such datasets. The repeated measures of an individual introduces temporal correlation between observations that must be controlled for. In addition, imaging studies frequently take multiple outcome measures from a single image which are spatially correlated.

To control for these two sources of correlation, George and Aban (2015) previously proposed the use of a linear model with a separable parametric spatiotemporal correlation structure. They showed that information criteria are highly accurate at choosing an appropriate combination of parametric spatial and temporal correlation functions that conserve the Type I error rate and maximize statistical power. George (2014) also demonstrated that such a model is better at conserving the Type I error rate and has higher power compared to certain summary methods (i.e., regional averages, endpoint analysis) which have previously been used to analyze longitudinal imaging studies [Ahmed et al. (2012), Schiros et al. (2012)].

The goal of this paper is to demonstrate how such a model would be applied to real-world longitudinal cardiac imaging data. The demonstration includes how a separable correlation structure can be chosen and evaluated, what kinds of inferences can be made using this model, and how it may give different results than summary methods. Note that this analysis strategy could be applied to any dataset with observations taken at multiple points in time and at multiple locations on or in a subject's body. Our application is a situation with a small number of discrete observations in time (5) and space (16), though we will discuss possible future extensions to larger datasets in the discussion.

2. Clinical application.

2.1. *Cardiac imaging and mitral regurgitation morphology.* The phrase *cardiac imaging* covers a wide variety of medical applications, from perfusion of oxygenated blood [Bowman and Waller (2004)] to the mechanical properties of the myocardium [Seals et al. (2014)] to the structure of the left ventricle (as in Figure 1) that we will consider in this paper. Most analyses of the left ventricle via cardiac MRI scans in recent years have been based on the spatial model proposed by the American Heart Association (AHA) [Cerqueria et al. (2002)]. This model sections the left ventricle into 17 segments, arranged as a cylinder with a hemispherical cap to represent the "bullet" shape of a healthy left ventricle (LV). The small number of segments is fairly specific to cardiology, with 17 being chosen as a compromise between the thousands of voxels available from a MRI scan and the rough spatial resolution of the gross anatomy (level from base to mid and side
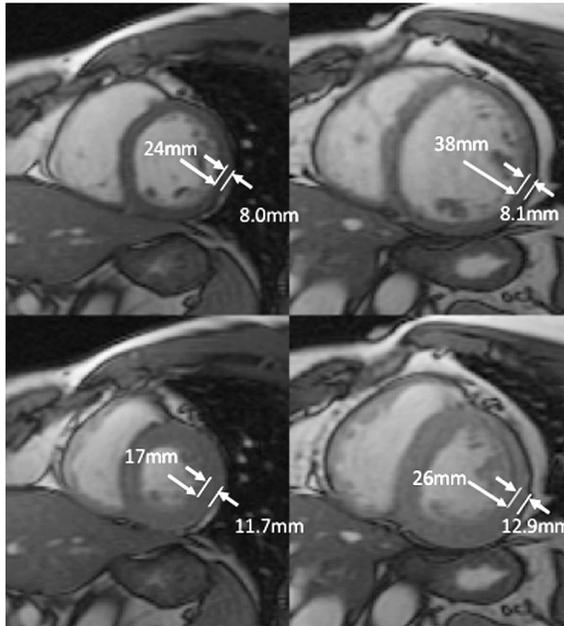
FIG. 1. *Structural cardiac MRI images of the left ventricle along the short axis for a healthy subject (*left*) and patient with mitral regurgitation (*right*). The top row shows the heart near k (*filling stage*) and the bottom row shows near end-systole (*contracting stage*). Note that for the MR case the radius of the left ventricle is larger while the wall thickness is similar, such that the end-diastolic radius-to-wall thickness (*R/T*) ratio at the marked point for the healthy subject is $\frac{24\ mm}{8\ mm} = 3$, while the MR patient's ratio is $\frac{38\ mm}{8.1\ mm} = 4.7$.*

of the heart) that practicing cardiologists are interested in interpreting. In practice, outcomes of interest such as wall thickness are averaged within each segment and taken as that segment's value. This may result in a loss of information for voxel-level outcomes, but for some mechanical properties (such as those we consider in this paper) that must be measured at a spatial resolution above the voxel this segmented model provides a convenient guide for summarization.

Our slightly modified version of the AHA model is given in Figure 2. Segment 17 at the tip of the apex was excluded, giving us outcomes measured at 16 spatial locations from a given imaging session. The 16 segment model was fit to a unit circle with the intersegment distances being the Euclidean distance between the centroids of each segment, shown as dots in Figure 2 and quantified in Table 1. The model is arranged as concentric circles representing the levels of the LV from the *base* (outside) to the *mid* then the *apex* (inside), and within these levels are four or six segments arranged circumferentially. The figure also denotes which primary coronary artery supplied blood to which segments.

In this paper we will consider the results of a trial [Ahmed et al. (2012)] looking at the effects of medical therapy on patients with chronic degenerative mitral regur-
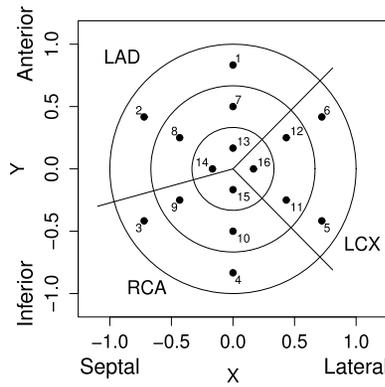
FIG. 2. *Plot of the* 16 *segments of the left ventricle. The outer ring corresponds to the base*, *the middle ring to the mid*, *and the inner circle to the apex* [*adapted from Cerqueria et al. (2002), Seals et al. (2014), George and Aban (2015)*]. *The right region corresponds to the segments supplied by the left circumflex* (*LCX*), *the upper left those supplied by the left anterior descending* (*LAD*), *and the bottom left those supplied by the right coronary artery* (*RCA*). *The numbers correspond to the segment's index as defined in Table* 1.

gitation (MR) with a primary outcome of the radius of curvature-to-wall thickness ratio (R/T ratio) measured repeatedly at multiple spatial locations. The R/T ratio is meant to be a measure of the sphericity of the left ventricle and the stress in the ventricular wall, which is relevant to mitral regurgitation as part of the natural progression of MR is the left ventricle becoming more spherical (Figure 1, right) rather than the healthy "bullet" shaped ellipsoid (Figure 1, left). As seen in Figure 1, increased sphericity (larger radius, smaller thickness) leads to a larger R/T ratio. Beyar et al. (1993) have shown that in healthy mammalian hearts the R/T ratio is approximately constant from the base to the apex of the left ventricle. Thus, the R/T ratio may be a better indicator of departure from the normal ventric-

TABLE 1
Spatial coordinates of the 16 *segments in the model of the left ventricle* [*Seals et al. (2014)*].
They are denoted as their level (*base*, *mid*, *apex*), *orientation*
(*anterior*, *septal*, *inferior*, *lateral*) *and index number*

Base, Ant. (1)	$(0, \frac{5}{6})$	Mid, Ant. (7)	$(0, \frac{1}{2})$	Apex, Ant. (13)	$(0, \frac{1}{6})$
Base, Ant. Sep. (2)	$(\frac{-5\sqrt{3}}{12}, \frac{5}{12})$	Mid, Ant. Sep. (8)	$(\frac{-\sqrt{3}}{4}, \frac{1}{4})$	Apex, Sep. (14)	$(\frac{-1}{6}, 0)$
Base, Inf. Sep. (3)	$(\frac{-5\sqrt{3}}{12}, \frac{-5}{12})$	Mid, Inf. Sep. (9)	$(\frac{-\sqrt{3}}{4}, \frac{-1}{4})$	Apex, Inf. (15)	$(0, \frac{-1}{6})$
Base, Inf. (4)	$(0, \frac{-5}{6})$	Mid, Inf. (10)	$(0, \frac{-1}{2})$	Apex, Lat. (16)	$(\frac{1}{6}, 0)$
Base, Inf. Lat. (5)	$(\frac{5\sqrt{3}}{12}, \frac{-5}{12})$	Mid, Inf. Lat. (11)	$(\frac{\sqrt{3}}{4}, \frac{-1}{4})$		
Base, Ant. Lat. (6)	$(\frac{5\sqrt{3}}{12}, \frac{5}{12})$	Mid, Ant. Lat. (12)	$(\frac{\sqrt{3}}{4}, \frac{1}{4})$		

ular structure, as it does not vary between segments to the extent of wall thickness or radius of curvature alone. The R/T ratio had been previously considered as an outcome in MR patients by Schiros et al. (2012), but only at the level of a global average or level-based averages analyzed separately, while this paper considers it for multiple segments with spatial correlation explicitly modeled. Considering the data lost when going to a single outcome value or the potential inflation in the Type I error rate when analyzing multiple locations without correction [George (2014)], our application of a 16 segment model constitutes an increase in the spatial resolution of the analysis. As suggested by Standaert, it is our hope that this increased resolution for segment-level outcomes or covariates grants new insights into the pathology of cardiac diseases.

2.2. *Clinical trial design and data structure.* The data used in this paper comes from the UAB SCCOR (Specialized Centers of Clinically Oriented Research) study, specifically Project 1-Aim 1, reported by Ahmed et al. (2012). This study was a randomized controlled phase IIb trial for the use of Toprol, a beta-blocker, in the treatment of patients with chronic degenerative MR. Beta-blockers were considered due to evidence of an elevated adrenergic response in MR patients [Nagatsu et al. (1994)] and how hyperactivation of the β-adrenergic pathway leads to a decrease in viability in myocardial cells [Mann et al. (1992)], possibly via oxidative stress [Ahmed et al. (2010)]. Supported by promising results in canine models [Tsutsui et al. (1994)], the intent was that treatment with β-blockers would prevent the stress and subsequent left ventricular dysfunction typically seen in MR patients. Thus, we shall test whether the treatment has an effect on the R/T ratio over time, such that a clinically important finding would be that it significantly reduces the increase in sphericity over time relative to the placebo.

In order for MR patients to be eligible they had to have moderate to severe mitral regurgitation characterized by mitral valve prolapse and thickening of its leaflets (assessed by an echocardiograph), left ventricular end-systolic dimension under 40 mm and left ventricular ejection fraction over 55%. Exclusion criteria included heart failure, prior myocardial infarction, coronary artery disease, kidney failure, hypertension and other valvular disorders. In other words, the patient cohort had normal cardiovascular health with the exception of having advanced mitral regurgitation. The study recruited 38 patients who were block randomized (in blocks of size 2 or 4) to either the treatment or control arm (19 per group). The baseline characteristics of the cohort are detailed in Ahmed et al. (2012) where the study was first reported, but a brief version is that the two treatment groups were balanced in size and did not significantly differ in demographics (age, sex, race) or baseline MRI-derived outcomes (end-diastolic volume, ejection fraction, peak early filling rate) and physical exam findings (blood pressure, pulse, New York Heart Association class).

After randomization, patients were dosed daily with either Toprol XL (a β_1-adrenergic receptor blocker) or the placebo and followed for two years.

Per protocol, cardiac MRI scans were taken at baseline and every six months af-
ter, giving cardiac imaging data for five discrete time points. The 3D MRI scans
gave information regarding the geometry and structure of the myocardium and
wall stress, and employed tissue tagging to quantify functional parameters such as
maximum strain in the left ventricular wall. Note that in this context, functional
refers to the kinematic properties of the myocardium such as stress, strain and rota-
tion; this study did *not* utilize myocardial perfusion imaging. The image was then
mapped to the standard 17-segment AHA mode [Cerqueria et al. (2002)], where
structural and functional outcomes were taken by averaging over each segment.
As mentioned above, segment 17 was excluded and the remaining 16 segments
were fit to a unit circle with the intersegment distances being the Euclidean dis-
tance between the centroids of each segment, shown in Figure 2 and quantified in
Table 1. Considering both spatial and temporal points, in the complete case each
subject had 80 observations (16 spatial observations at each of 5 time points). The
mean R/T ratio for each group at each segment and time point is given in Figure 3;
note that this figure is meant to be descriptive of the time courses for each segment
and that the error bars should not be used for inference, as doing 80 simultaneous
correlated tests without correction is statistically unsound. In this paper we have
chosen to look at the end-diastolic R/T ratio, although as one can see in Figure 1
the end-systolic values would also show differences in sphericity.

    In this analysis, we initially considered the 38 randomized subjects who had
longitudinal imaging data: 29 subjects had complete data with the other 9 subjects
having missed one or more follow-up visits (3 on Toprol, 2 on placebo) or attended
but were missing some MRI data (2 Toprol, 2 placebo). A total of 9 follow-up visits
were missed for a loss-to-follow-up rate of 5%. In addition, 8 subjects (6 placebo,
2 Toprol) had mitral valve repair surgery during the study; two of the placebo
group underwent surgery immediately following randomization but had all five
visits recorded as part of a separate arm of the SCCOR study. Note that the study
treatment was stopped following surgery. The initial analysis was performed using
an intent-to-treat design that included all of the observations from subjects under-
going surgery (including the two who had immediate surgery), but a secondary
sensitivity analysis was performed with those subjects' post-surgery observations
excluded from the analysis, as the patients were removed from Toprol or placebo
after surgery. The results of the sensitivity analysis are given in the supplemental
article [George et al. (2016)]. Along the lines of intent-to-treat, the planned visit
times were used in the analysis as the time for the subjects' visits.

3. Statistical model.

3.1. *Linear model with a separable parametric correlation structure.*   To ana-
lyze the longitudinal imaging data, we look to use a linear model with a separable
parametric correlation structure. The theoretical and technical details have been
described in our previous work [George and Aban (2015), George (2014)], but we
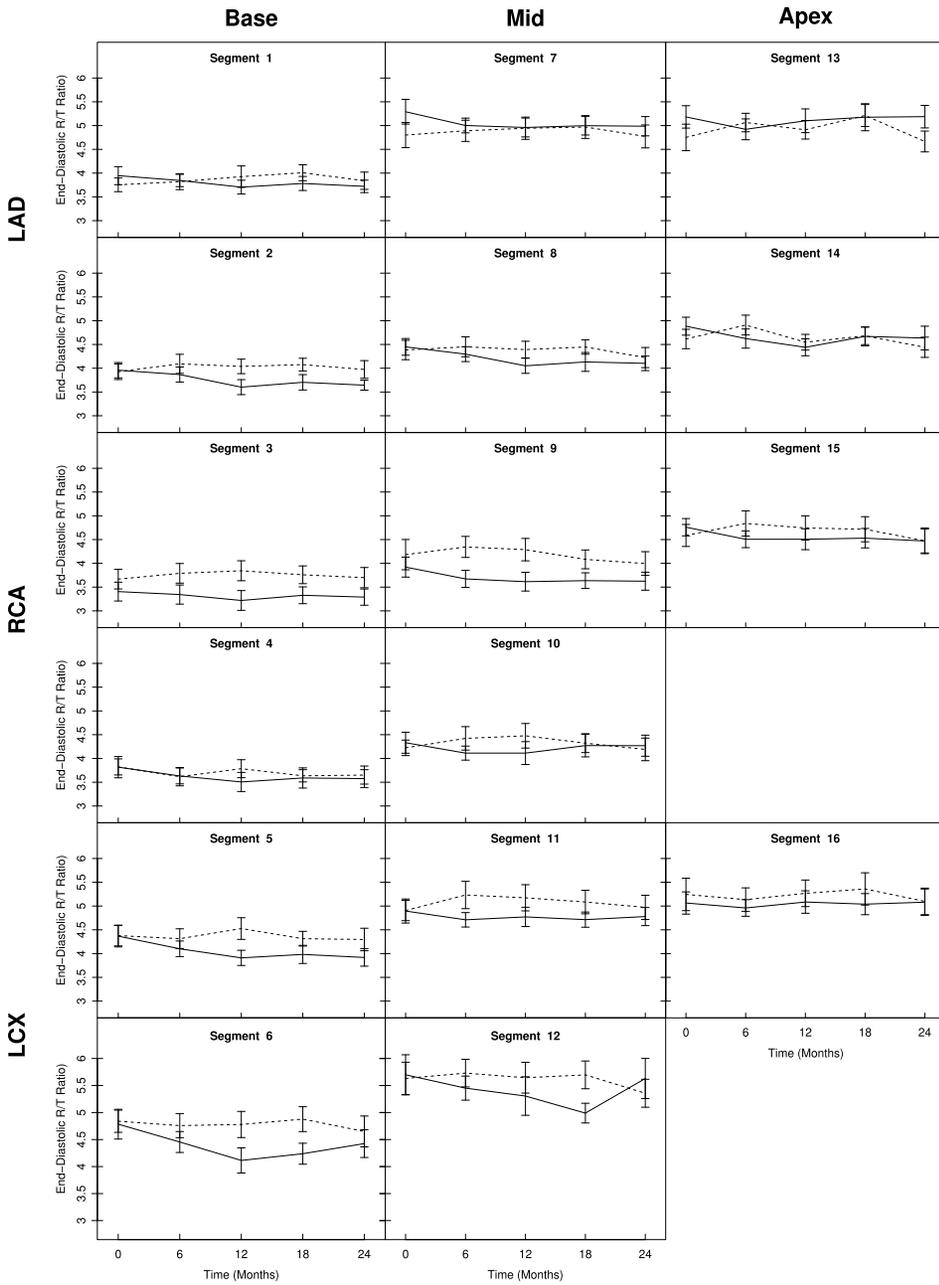
FIG. 3. *Plot of the time courses of the mean end-diastolic R/T ratios among all subjects in each group at the* 16 *segments of the left ventricle. The solid lines represent the average time courses of the placebo group, and the dashed line the courses of the treatment group. The error bars demark one standard error of the mean in each direction* (*total width is two standard errors*) *and are shown for the purpose of describing the spread of the R/T ratio among subjects and should not be taken as formal inference.*

shall discuss the parts of the model here. The base is a *linear model*, defined in equation (1), where the mean response is a sum of predictor variables ($\mathbf{X}_i$) and their associated coefficients ($\boldsymbol{\beta}$):

$$(1) \qquad \mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i, \qquad \boldsymbol{\varepsilon}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma}_i), \qquad \boldsymbol{\Sigma}_i = \sigma^2(\boldsymbol{\Sigma}_T \otimes \boldsymbol{\Sigma}_S).$$

This structure allows for the use of subject-specific predictors (i.e., age, sex, race), time-varying predictors (i.e., systolic blood pressure at each visit), space-varying predictors (such as what part of the heart an outcome is from) and linear (or higher order) trends over time or space. We assume a multivariate normal distribution with constant variance for a subject's responses.

The other parts refer to how the spatial and temporal correlation is modeled. Unlike simple linear regression, where all of the observations are independent and uncorrelated, our model directly quantifies the correlation between the outcome measurements. The *separable* nature of the correlation structure means that the spatial and temporal correlation can be handled separately so that how correlation changes over space does not influence how it changes over time. Mathematically, this is done by taking the Kronecker product ($\otimes$) of temporal ($\boldsymbol{\Sigma}_T$) and spatial ($\boldsymbol{\Sigma}_S$) correlation matrices. The assumption of separability not only makes interpretation easy but also makes the math simpler as well: the correlation between two observations is the product of their spatial and temporal correlations.

Parametric refers to how the correlation over space or time is modeled as a parametric function. We define $\boldsymbol{\theta}$ as the vector of parameters in $\boldsymbol{\Sigma}_i$. For example, an exponential or autoregressive-1 function implies that the correlation decreases exponentially at a constant rate over space or time, respectively. The benefit of parametric functions is that they model the correlation between a large number of observations while estimating a small number of parameters. This is in contrast to an unstructured correlation model whose parameters increase with the square of the number of locations; 16 spatial locations results in an unstructured matrix with 120 parameters, which may be inefficient (if not impossible) for an imaging study where the number of subjects is often limited. If the number of repeated measures is small, such as for a limited number of follow-up visits, then the unstructured correlation model may be viable.

Estimation of the model is best done using restricted maximum likelihood estimation (REML), as discussed by George and Aban. Unfortunately, the estimation of a linear model with the kind of separable parametric correlation structure we proposed is not currently supported by common statistical software such as SAS (v9.4) [SAS Institute Inc. (2014)] or R (lme4 package v1.1-7) [Bates et al. (2014)]. Therefore, the estimation would need to be done by hand coding an estimation algorithm [as done by Simpson et al. (2014a) in SAS PROC IML] or by utilizing specialized commercial statistical software such as ASReml [Butler et al. (2007)].

One challenge in implementing this model is the selection of parametric functions for the correlation structure. In some ways this can be seen as needing to be

done twice since a spatial function and a temporal function must be chosen, but it is best to examine many potential combinations concurrently. In practice, it may be necessary to fit a large number of combinations to be comfortable that one models the correlation well. Numerous resources exist that define spatial [Waller and Gotway (2004)] and temporal [Littell, Pendergast and Natarajan (2000); Schaalje, McBride and Fellingham (2002); Vallejo et al. (2011)] correlation functions. It is often but not always the case that functions with a larger number of parameters offer a better fit to the observed correlation. In general, one wishes to balance the number of parameters in $\boldsymbol{\theta}$ with the goodness of fit of the model. To this end, George and Aban found that, in the scenarios considered, information criteria (AIC, AICC, BIC, CAIC and HQIC) were highly (over 90% in some cases) accurate at choosing the true correlation structure, and that at the least they reliably choose a structure that will conserve the Type I error rate and maximize power when given a sufficient number of structures to choose from. Information criteria are simple to use, as each structure is given a score and the one with the smallest score is the "best" model. Although there are a large number of information criteria that one could use, the most popular two (AIC and BIC) appear in most statistical programs; note that of the two the BIC may be much more accurate at separable correlation structure selection than the AIC [George and Aban (2015)].

3.2. *Summary methods.*   Spatial summary measures have been previously used in the analysis of structural cardiac imaging data [Ahmed et al. (2012), Schiros et al. (2012)]. Although the use of summarization methods may seem anathema to statisticians experienced in analyzing voxel-level data as this approach "throws out" information, there are several reasons for its prior use in cardiology. The outcomes of interest are related to the overall morphology of the heart; an analysis with an extremely high spatial resolution is not necessarily more useful to the clinician than a rougher one [Cerqueria et al. (2002)]. Due to the highly interdependent nature of the myocardium with regard to the dynamics of the heart, structural and functional outcomes are expected to have high spatial correlation between segments. This means that ignoring spatial correlation can have deleterious effects on the validity of statistical inference [George (2014)]. It also means that there is not as much unique information among the segments, so not as much actual information is lost in the summarization. Last, there have only been a small number of studies [Bowman and Waller (2004), Seals et al. (2014)] that have looked to introduce spatial analysis to cardiac imaging data, so it has not yet become the standard research approach in the field.

The spatial summary methods considered here include taking a *global average* of all of the observed segments in the left ventricle and *regional averages* of the three levels of the left ventricle. The global average eliminates the spatial correlations between observations in the analysis but loses the most information. For regional averages, the number of regions is typically chosen to retain some spatial resolution while still being able to estimate an unstructured correlation matrix

between the regions. Taking averages within the three levels results in an unstructured matrix with only three parameters, a reasonable amount. Note that regional averages should be analyzed together while modeling the correlation between the regions; failure to do so by analyzing the regions separately can result in extreme inflation of the Type I error rate due to multiple testing issues [George (2014)]. Care must also be taken in the definition of the regions, as redefining the boundaries can drastically alter the results of an analysis due to the modifiable area unit problem [Waller and Gotway (2004)]. In addition to the potential loss of information, another drawback to spatial summary methods is the inability to consider predictors or covariates that cannot be defined at the global or regional level. We will explore this further in the applied example.

For completeness, we will also consider the use of the temporal summary method of *endpoint analysis*. Endpoint analysis, also called change score analysis, considers the pre-post change in the outcome and is commonly used in randomized controlled trials. A notable benefit of endpoint analysis is that it removes the need to model the specific shape (i.e., linear or nonlinear) of the time course of the outcome when the overall change over time is of interest [Matthews (1990)]. Of course, this approach may have lower statistical power if more than one follow-up observation is available. A temporal summary method such as this obviates the need to the modeling of temporal correlation; this benefit may not be substantial by itself given the ubiquity of longitudinal data analysis methods, but it allows for a spatial correlation model to be used to analyze spatiotemporal data without a joint correlation structure using mainstream (e.g., SAS, R) software.

Although the use of spatial and/or temporal summary methods would seem to eliminate the explicit separability assumption, this is not necessarily the case. Unless a heterogeneous variance model is used or both methods are used to get a single outcome value per subject, nonseparability of spatial and temporal correlation would result in heteroscedasticity among temporal summarizations between segments or spatial averages between time points. As these summary methods are at their core weighted averages, heteroscedasticity may also arise between subjects in the presence of spatially or temporally missing data, as the averages would be the result of different quantities of observed values if imputation procedures are not used [Everitt (1995), Fitzmaurice, Laird and Ware (2004)]. Improper handling of either source of heteroscedasticity can result in an inflated Type I error rate [Gibbons et al. (1988), Frison and Pocock (1992)].

4. Statistical analysis.

4.1. *Linear model with a separable parametric correlation structure.*   To compare statistical methodologies, we analyzed the SCCOR data with our proposed model as well as with a battery of summary measures. Our linear model has the form described in equation (1) where $\mathbf{X}_i \boldsymbol{\beta}$ are predictor variables and their associated parameters and $\mathbf{Y}_i$ are the observed end-diastolic R/T ratios for subject i.

Note that due to observed skewness of the model residuals, a log-transformation of the R/T ratio was used (see the supplemental article for QQ plots [George et al. (2016)]). In our application, we are looking to fit the model

$$
\begin{aligned}
\mathbf{X}_i \boldsymbol{\beta} = {} & \beta_0 + \beta_1 \text{Sex}_i + \beta_2 \text{Group}_i + \beta_3 \text{Time}_{ij} + (\beta_4 \text{Mid}_k + \beta_5 \text{Apex}_k) \\
& + (\beta_6 \text{RCA}_k + \beta_7 \text{LCX}_k) + \beta_8 \text{Time}_{ij} \text{Sex}_i + \beta_9 \text{Time}_{ij} * \text{Group}_i \\
& + (\beta_{10} \text{Time}_{ij} * \text{Mid}_k + \beta_{11} \text{Time}_{ij} * \text{Apex}_k) \\
& + (\beta_{12} \text{Time}_{ij} * \text{RCA}_k + \beta_{13} \text{Time}_{ij} * \text{LCX}_k),
\end{aligned}
$$

(2)

where Time_{ij} was the time of subject i's jth observation; Sex_i was an indicator variable for whether the subject was male ($=1$) or female ($=0$); Group_i was an indicator variable for the treatment group for subject i, be it placebo ($=0$) or beta-blocker ($=1$); Mid_k and Apex_k were indicator variables for whether the ijkth observation was from the mid or apex of the left ventricle, respectively, with the base as the reference group; RCA_k and LCX_k were indicator variables for whether the ijkth observation was from the segment supplied by the right coronary artery (RCA) or the left circumflex (LCX), respectively, with the left anterior descending (LAD) as the reference group. Note that the coronary artery designation should *not* be interpreted as cardiac perfusion necessarily being related to ventricular remodeling in mitral regurgitation patients; we simply use it as a convenient way to spatially subdivide the left ventricle circumferentially, as the six segments in the base and mid do not neatly subdivide into nonoverlapping anterior/inferior and lateral/septal groups. Thus, dividing the segments into three circumferential regions allows us to test differences between the sides, as the radius and wall thickness are known to vary circumferentially. The parentheses in equation (2) denote terms corresponding to the level of the left ventricle or the side of the heart, such that inferences made about the terms inside are done together in a two degree of freedom test. A quadratic time effect was considered but dropped, as it was not found to be significant. Of particular interest is the treatment-by-time interaction, as it would indicate a treatment effect on the left ventricular remodeling in these patients.

Our model also assumes that $\boldsymbol{\varepsilon}_i$ follows a multivariate normal distribution with mean zero and a separable parametric correlation structure. For the correlation structure, we considered all twelve combinations of three spatial correlation functions (exponential, spherical and Matérn) crossed with four temporal correlation functions (compound symmetric [CS], autoregressive-1 [AR-1], Toeplitz and unstructured [UN]). We chose a working correlation structure via BIC with a sample size adjustment of the total number of subjects in the dataset, which George and Aban (2015) found to be reliably accurate at choosing the true correlation structure. To confirm our choice, we also plotted the estimated spatial and temporal correlation functions versus the observed correlation between pairs of observations. To calculate the 120 unstructured spatial correlation parameters, we fit the

model on pairs of segments from all subjects at all observed time points with an
unstructured temporal correlation model.

The linear models were fit using the ASReml-R package (v. 3.0, VSN Interna-
tional, Hemel Hempstead, UK) [Butler et al. (2007)]. The model with the chosen
covariance structure then had its fixed effects tested with a conditional F-test with
a Kenward–Roger adjustment for the denominator degrees of freedom at an α-level
of 0.05 [Kenward and Roger (1997)]. The corrected F-test is generally considered
to be better to use than a simple Wald's test when there are correlation parameters
in the model and the sample size is small; this assertion is supported by simula-
tion studies by George (2014) which found the corrected F-test to conserve the
Type I error rate without a meaningful loss in power for sample sizes compara-
ble to typical longitudinal imaging studies. Missing observations were handled by
using the full information likelihood approach [Allison (2012)]. The code used in
this analysis is available from the corresponding author by request.

4.2. *Summary methods.* We implemented the summary measures in space and
time discussed in Section 3.2 to demonstrate the benefits of our model compared
to previous methods for analyzing longitudinal imaging data. The spatial summary
measures considered included a global average of all 16 segments and averages
with the levels (base, mid and apex) analyzed jointly, resulting in one and three
spatial observations per subject per time, respectively. As with the spatiotemporal
model, missing data was handled using the full information likelihood. On the
temporal side, we considered endpoint analysis of the change between the baseline
and two-year observations, and excluded the subject from analysis if either of those
two observations were missing. Note that when the temporal summary measures
were used, the fixed effects relate whether a predictor (such as treatment group)
affects the overall change over time. All combinations of these summary measures
along with the direct modeling of spatial and temporal correlation led to a total of
six models to compare. When one of the models needed a parametric correlation
function chosen, the BIC was used. Convergence of the model estimation was only
an issue for the level averages as the inter-level correlations were near 1; these
problems were overcome by setting initial values of the correlation parameters to
be the Pearson correlation estimates between levels.

The fixed effects used in the six models are detailed in Table 2, and were chosen
for uniformity in what the models controlled for. All of the models allowed for the
use of subject-level predictors such as the treatment group and the subject's sex.
Sex was chosen as it may relate to the geometry of the heart through a subject's
size or shape. The summary measures may restrict the use of other predictors,
as mentioned above. The level averages prevent the use of space-varying covari-
ates besides those identifiable at the subject level; in our example the side of the
heart cannot be modeled when the level averages are used. The temporal summary
method similarly prevents the use of time-varying covariates or even the use of the
subject's exact visit time as a predictor.

TABLE 2

*Six linear models made from combinations of spatial and temporal methods, the form of their fitted predictors, and what predictor (denoted in **bold**) is the focus of hypothesis testing for whether the treatment* (Trt$_i$) *changes the time-course of disease progression*

Spatial method	Temporal method	Fitted model
Correlation	Correlation	$E[Y_{ijk}] \sim 1 + \text{Sex}_i + \text{Time}_{ij} + \text{Trt}_i + \text{Mid}_k + \text{Apex}_k$ $+ \text{RCA}_k + \text{LCX}_k + \text{Time}_{ij}\text{Sex}_i + \textbf{Time}_{ij}\textbf{Trt}_i$ $+ \text{Time}_{ij}(\text{Mid}_k + \text{Apex}_k) + \text{Time}_{ij}(\text{RCA}_k + \text{LCX}_k)$
	Endpoint	$E[Y_{ik}] \sim 1 + \text{Sex}_i + \textbf{Trt}_i + \text{Mid}_k + \text{Apex}_k + \text{RCA}_k + \text{LCX}_k$
Level average, together	Correlation	$E[Y_{ijm}] \sim 1 + \text{Sex}_i + \text{Trt}_i + \text{Time}_{ij} + \text{Mid}_m + \text{Apex}_m$ $+ \text{Time}_{ij}\text{Sex}_i + \textbf{Time}_{ij}\textbf{Trt}_i + \text{Time}_{ij}(\text{Mid}_m + \text{Apex}_m)$
	Endpoint	$E[Y_{im}] \sim 1 + \text{Sex}_i + \textbf{Trt}_i + \text{Mid}_m + \text{Apex}_m$
Global average	Correlation	$E[Y_{ij}] \sim 1 + \text{Sex}_i + \text{Trt}_i + \text{Time}_{ij} + \textbf{Time}_{ij}\textbf{Trt}_i + \text{Time}_{ij}\text{Sex}_i$
	Endpoint	$E[Y_i] \sim 1 + \text{Sex}_i + \textbf{Trt}_i$

5. Results.

5.1. *Linear model with a parametric spatiotemporal covariance structure.* Let us first focus on the application of our previously proposed spatiotemporal model. The first step when using it is to choose appropriate spatial and temporal correlation functions. Using the BIC to pick a model, we found that a Matérn-by-unstructured correlation model provided the "best" balance of goodness of fit and simplicity for the observed data (Table 3). This was the most complex correlation structure we considered, so it warranted investigation of how well the model truly fit the data.

Figure 4 shows the estimated correlation functions along with the unstructured correlation estimates between time points and segments. The unstructured temporal correlation provides the best possible fit, and we can see that the simpler parametric functions do not necessarily fit the observed correlation very well. We observed that the unstructured spatial correlation seemed to have a random scatter when plotted versus the distance between segments, which would make parametric modeling extremely difficult. However, of the three functions considered, it did seem that Matérn gave the best fit, as the others more severely underestimated the true correlation between far apart segments. For completeness, a compound symmetric model was tried, but its associated BIC (-9313.4) was not superior to a Matérn function, suggesting there is indeed a slight downward trend of correlation over distance. Note that this approach of graphically examining the fits of the estimated correlation functions can also be highly useful when multiple information criteria provide conflicting answers for which model is "best."

*Table of the BIC for each of the twelve fitted correlation structures, for the datasets with all
observations or with the follow-up visits post-surgery excluded from the model.
The smallest value in each column is the chosen model via BIC and is denoted in **bold***

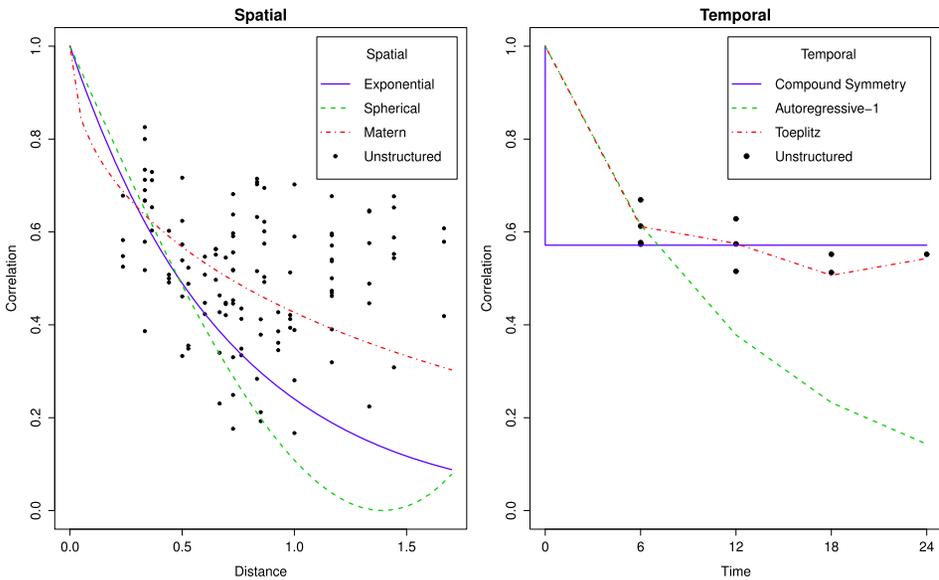| Correlation structure	Length of θ	BIC	
		All data	**Post-surgery excluded**
EXP ⊗ CS	3	−9433.7	−7967.0
SPH ⊗ CS	3	−9234.8	−7799.8
MAT ⊗ CS	4	−9505.5	−8027.1
EXP ⊗ AR-1	3	−9250.0	−7837.5
SPH ⊗ AR-1	3	−9062.5	−7681.9
MAT ⊗ AR-1	4	−9305.2	−7880.9
EXP ⊗ TOE	6	−9467.3	−7994.8
SPH ⊗ TOE	6	−9269.3	−7828.6
MAT ⊗ TOE	7	−9537.1	−8052.6
EXP ⊗ UN	12	−9480.6	−8013.1
SPH ⊗ UN	12	−9285.1	−7850.2
MAT ⊗ UN	13	**−9544.1**	**−8063.3**

FIG. 4.   *Plot of the unstructured spatial and temporal correlations and the associated estimated
correlation functions for the natural log of the end-diastolic R/T ratio using the SCCOR data includ-
ing post-surgery observations. On the left are the spatial functions when the temporal correlation is
unstructured, while on the right are the temporal structures for a Matérn spatial structure.*

TABLE 4
The parameter estimates and associated statistical inference for each of the predictors in the model
for the natural log of the R/T ratio with a Matérn-by-unstructured correlation matrix,
including post-surgery observations

Predictor	Parameter estimate	F	p-value
Intercept	1.44647	$F_{1,162} = 4864.00$	<0.0001
Sex (Male)	−0.18809	$F_{1,162} = 22.30$	<0.0001
Time	−0.00313	$F_{1,159.9} = 2.24$	0.1365
Treatment	0.01022	$F_{1,162} = 0.92$	0.3398
Mid	0.15700	$F_{2,532.5} = 152.50$	<0.0001
Apex	0.20666		
RCA	−0.06461	$F_{2,566.6} = 56.62$	< 0.0001
LCX	0.07954		
Sex ∗ Time	0.00182	$F_{1,159.9} = 1.12$	0.2917
Treatment ∗ Time	0.00217	$F_{1,159.9} = 1.60$	0.2073
Mid ∗ Time	0.00012	$F_{2,548.1} = 0.07$	0.9337
Apex ∗ Time	−0.00005		
RCA ∗ Time	−0.00038	$F_{2,550.5} = 0.19$	0.8277
LCX ∗ Time	−0.00009		

After the correlation structure has been chosen and the assumption of normality has been checked, the next step is to review the estimates and inferences about the predictors in the model. The estimates of and inference on the parameters given in equation (2) are given in Table 4. As mentioned, we used a F-test with a Kenward–Roger correction for the denominator degrees of freedom; the effect of the correction can be seen in the reported test statistics, as the denominators are not integers.

The parameter of interest, the treatment-by-time interaction β_9, was not significantly different from zero ($p = 0.2073$). This means that there is no evidence that treatment with beta-blockers affects the remodeling in patients with mild mitral regurgitation. The linear trend over time was also not significant ($p = 0.1365$), which suggests that the R/T ratio is simply not changing very much over the twenty-four months of observation. The effect of sex was strongly significant ($p < 0.0001$), with men having slightly smaller R/T ratios than women; this is likely due to men being larger and having larger hearts such that the larger wall thickness overrides the larger radius of curvature. The level of the left ventricle was also highly significant ($p < 0.0001$), with the R/T ratio increasing from the base down to the apex. We expected to see this trend, as it matches what is seen in the natural progression of left ventricular remodeling due to mitral regurgitation. The side of the left ventricle was also significantly associated with the R/T ratio ($p < 0.0001$) such that the lateral side (LCX) had the highest ratio, followed by the anterior/septal side (LAD), then the inferior/septal side (RCA). It is possible that this trend is due to

the balancing forces from other chambers of the heart, especially the right ventricle opposite the septum, that are lacking in the lateral side which only has the pericardium restraining the myocardium. Although the pericardium is not elastic and typically acts to prevent ballooning of the left ventricle, it can weaken and stretch over time with chronic pressure from an overloaded left ventricle. Such unbalanced remodeling has been previously seen by Young et al. (1996) in canines with induced mitral regurgitation. However, there is a natural difference in the R/T ratio between the septal and lateral ventricular walls seen in healthy patients that may be the source of this statistically significant difference. Assuming the R/T ratio is a valid measure of sphericity, this may suggest that the highest sphericity in mitral regurgitation patients is in their left ventricles' lateral and apical region. These results match what was seen in Figures 1 and 3.

The sensitivity of the intent-to-treat analysis strategy was considered by a secondary analysis where all observations taken after valve repair surgery were excluded from the dataset. The results of the analysis did not noticeably change, as the BIC still chose a Matérn-by-unstructured model (Table 3) and none of the inferences changed at a $\alpha = 0.05$ level (see the supplemental article [George et al. (2016)]). The estimates themselves changed slightly, suggesting that surgery may not be independent of the R/T ratio and the estimates should be interpreted cautiously.

5.2. *Summary methods.*   Now that we have considered the results of applying our proposed model to the UAB SCCOR data, it is of interest to compare it to the results from summary methods commonly used in longitudinal imaging data analysis. For the sake of brevity, we shall only consider inference upon the effect of medical therapy on the time course of the end-diastolic R/T ratio, log-transformed to correct for skewness. The test statistics for the corrected F-test and associated p-values for the six models are given in Table 5 along with what correlation structure was chosen (if applicable) via BIC to fit the data.

The first thing to note is that the qualitative conclusions did not change: all p-values are greater than 0.05. We can also note that the denominator degrees of freedom are the largest for our model and much smaller for the summary methods. In every case the corrected degrees of freedom are smaller than the number of observations used in the analysis, although the difference varies with the summary measure used. The spatial summary measures only slightly reduced the degrees of freedom despite a large reduction in the number of observations, while endpoint analysis reduced the degrees of freedom to around the number of independent subjects. This reduction in the degrees of freedom is expected in a Kenward–Roger adjustment, and reflects how the information is condensed and possibly lost. It may also reflect how despite summary measures simplifying the dimension of correlation among the observations, they may increase the correlation between the remaining measures. For example, the correlations between the level averages were around 0.9 through the various approaches, which suggests that the information

TABLE 5
Inferences about a treatment effect over time on the log of the end-diastolic R/T ratio from the six combinations of spatial and temporal methods from the UAB SCCOR study using all observed outcomes

Spatial method	Temporal method	No. of obs.	Fitted Σ structure	Est. effect per month	Test statistic	*p*-value
Correlation	Correlation	2894	Matérn ⊗ UN	0.00217	$F_{1,159.9} = 1.60$	0.2073
	Endpoint	574	Matérn	0.00204	$F_{1,37} = 1.17$	0.2862
Level average	Correlation	543	UN ⊗ CS	0.00147	$F_{1,154.0} = 0.83$	0.3646
	Endpoint	108	UN	0.00151	$F_{1,32.9} = 0.57$	0.4674
Global average	Correlation	181	CS	0.00148	$F_{1,141.2} = 0.91$	0.3416
	Endpoint	36	N/A	0.00144	$F_{1,33} = 0.53$	0.4709

from those three observations is far less than three independent observations and closer to a single observation.

Comparing the different methods, it seems that the use of spatial correlation estimated a larger treatment-by-time effect than spatial summary measures, despite the target of inference not having a spatial component. It is possible that the ability to include space-varying covariates such as the coronary arteries affected this estimation. Endpoint analysis provided similar estimates to the temporal correlation model, though it appears to have reduced power; this is likely due to it inducing a higher proportion of missing data and having to use fewer observations.

The results of the summary method comparison were slightly different when the post-surgery observations were excluded, as seen in the supplemental article [George et al. (2016)]. All of the six methods still had *p*-values greater than 0.05, although the estimates changed slightly. The spatial comparisons did not change, as to be expected given the missingness was temporal rather than spatial. The temporal methods were affected differently, particularly in the number of observations used. The temporal correlation approach lost the relevant observations but still retained the six subjects' pre-surgery data. The nature of the cutoff meant that endpoint analysis had to count the eight as missing, dropping the number of included subjects from 36 to a mere 28. Some of these changes could be due to how the methods handle missing data, but there is also the concern regarding this type of missingness. Since the excluded observations are from subjects who underwent valve repair surgery and surgery is only done on patients whose mitral regurgitation has progressed far enough, then any measure of disease progression (such as the R/T ratio) could not be missing completely at random (MCAR) if post-surgery observations are excluded. The missingness also did not affect the two groups equally; six versus two may seem trivial, but when the original group sizes were nineteen and the missingness is not completely at random, the missingness should be carefully considered. However, since at least some data was retained on

these subjects, the inference would still be valid as long as the pattern is missing at random (MAR) versus not missing at random (NMAR), as our method can utilize the full information likelihood [Allison (2012)]. Fortunately, it is very possible that we would be under the MAR situation given the high temporal correlation between time points that could inform what later unobserved values would be.

It should also be noted that although in this study the qualitative results (all $p > 0.05$) did not change between the different spatiotemporal methods, the numbers themselves did. It would certainly be possible for significance to change between the methods when they are applied to a different dataset. As always, scientific justification should be used to choose a method instead of "cherry picking" the one that gives the most favorable p-value.

6. Discussion. In this paper we have described how a linear model with a separable parametric correlation structure could be used in practice, and have illustrated the method using data from a longitudinal imaging study. Only general guidelines can be given, as each application has its own nuances. A general strategy for implementing our proposed model on spatiotemporal data could be considered as such:

1. Decide on all of the predictors in the analysis that would be of interest to scientific investigators.
2. Decide on a number of spatial and temporal correlation functions to try to fit to the data. Functions with different properties should be considered, such as different shapes and a mixture of simpler functions and more complex functions which may have greater flexibility. This step should also be a collaboration between statistician and investigator, as the functions should be able to model the correlation behavior expected by prior scientific knowledge. In practice, this may result in fitting every model supported by the chosen software that is not scientifically unreasonable.
3. Fit linear models with all of those predictors included in the fixed effects for a wide variety of combinations of spatial and temporal correlation functions. If the number of combinations considered is too small, it is possible that none of them will model the correlation sufficiently well.
4. Choose between models using information criteria. One should also compare the estimated correlation to the observed correlation; graphical methods are highly useful to assess goodness of fit. If none of the fitted structures seem appropriate, additional approaches to modeling the covariance should be considered.
5. Perform inferences on the fixed effects using the model with the chosen correlation structure.
6. If some predictors are not significant, a more parsimonious model can be obtained with backward selection. Note that the above steps for choosing a correlation structure must be repeated for each new set of fixed effects.

The greatest challenge to this approach is finding a correlation structure that fits the data well. Properties of correlation functions that should be considered include whether correlation is strictly decreasing with distance, and the shape and rate of decay of the function. There has been an immense amount of work done to define valid parametric correlation functions (too many to list exhaustively here), so one option would be to simply try more structures, such as the flexible linear exponent autoregressive (LEAR) function [Simpson et al. (2014a)]. This may require statistical programming to augment or develop the model estimation software if the desired functions are not already supported. Another point that should be considered is the assumption of separability; if there is an interaction between spatial and temporal correlation, then no pairing of separate functions will properly model the true correlation. Much work has been done to test this assumption of separability, but a good starting point would be a recent likelihood ratio test proposed by Simpson et al. (2014b) that was designed with longitudinal imaging studies in mind. More statistical research needs to be done to determine how sensitive a model like ours is to violations of separability and what nonseparable methods are appropriate to use in our given application. The assumption of multivariate normality is also highly important and should be checked; deviations can possibly be helped by a transformation to the outcome values. In addition, as our spatiotemporal model is likelihood based, a Bayesian extension would be a logical next step.

As we have seen, summary methods provide a way to analyze longitudinal imaging data without correlations. Unfortunately, spatial summary methods may provide different estimates even if the predictor of interest is not space-varying. Furthermore, both spatial and temporal summary measures preclude the use of space- and time-varying covariates/predictors, respectively. In our example, we would not have been able to observe the association between the side of the heart and the R/T ratio had we used spatial summary measures. As such, summary measures should probably be avoided unless there is a highly specific scientific justification.

Another issue to consider when doing the analysis is how to handle the imperfections of real data. As our model estimation is done by maximizing the restricted likelihood, missing values can be handled well by using the full information likelihood provided the data is MCAR or MAR [Allison (2012)] and follows all of the caveats of using likelihood-based methods in the presence of missing data. Uneven follow-up times are more difficult, as they preclude the use of many temporal correlation functions that assume there are a finite number of evenly spaced observations. One option (which we have used in this paper) is to use the planned observation times; this allows the use of an unstructured temporal correlation model which is desirable but does involve ignoring information that was collected. One possibility would be to utilize correlation functions that are traditionally considered to be spatial, using the true observation time as the distance, but such an approach needs statistical validation before it can be recommended.

Last, we would like to discuss applications of our model in fields other than structural cardiac imaging. The largest areas would be in functional (e.g., perfusion) and neuroimaging, which we have avoided mentioning thus far to avoid confusion of our application in this paper. The most significant challenge is that such data has orders of magnitude greater numbers of spatial (hundreds of thousands of voxels) and temporal (multiple measures per second) observations. While this high dimension prevents a reasonable application of unstructured correlation matrices, further research is needed to determine if separable parametric spatiotemporal correlation structures would be appropriate for that application.

Acknowledgments. We would like to thank the reviewers whose comments improved the final manuscript.

SUPPLEMENTARY MATERIAL

Supplementary tables and figure (DOI: 10.1214/16-AOAS911SUPP; .pdf). We provide QQ plots of the model residuals justifying the log transformation, and tables of the regression parameters for the primary spatiotemporal model and inference about a treatment-by-time interaction when the post-surgery observations were excluded from the analysis.

REFERENCES

AHMED, M. I., GLADDEN, J. D., LITOVSKY, S. H., LLOYD, S. G., GUPTA, H., INUSAH, S., DENNY, T. JR., POWELL, P., MCGIFFIN, D. C. and DELL'ITALIA, L. J. (2010). Increased oxidative stress and cardiomyocyte myofibrillar degeneration in patients with chronic isolated mitral regurgitation and ejection fraction > 60%. *J. Am. Coll. Cardiol.* **55** 671–679.

AHMED, M. I., ABAN, I., LLOYD, S. G., GUPTA, H., HOWARD, G., INUSAH, S., PERI, K., ROBINSON, J., SMITH, P., MCGIFFIN, D. C., SCHIROS, C. G., DENNEY, T. S. and DELL'ITALIA, L. J. (2012). A randomized controlled phase IIb trial of beta-1-receptor blockade for chronic degenerative mitral regurgitation. *J. Am. Coll. Cardiol.* **60** 833–838.

ALLISON, P. D. (2012). Handling missing data by maximum likelihood. SAS Global Forum 2012, Paper 312-2012, Available at http://statisticalhorizons.com/wp-content/uploads/MissingDataByML.pdf.

BATES, D., MAECHLER, M., BOLKER, B. and WALKER, S. (2014). Package 'lme4' [Software]. Available at http://lme4.r-forge.r-project.org/.

BEYAR, R., WEISS, J. L., SHAPIRO, E. P., GRAVES, W. L., ROGERS, W. J. and WEISFELDT, M. L. (1993). Small apex-to-base heterogeneity in radius-to-thickness ratio by three-dimensional magnetic resonance imaging. *Am. J. Physiol.* **264** H133–H140.

BOWMAN, F. D. and WALLER, L. A. (2004). Modelling of cardiac imaging data with spatial correlation. *Stat. Med.* **23** 965–985.

BUTLER, D., CULLIS, B. R., GILMOUR, A. R. and GOGEL, B. J. (2007). ASReml-R reference manual (Release 2.00) [Software]. Available at http://www.vsni.co.uk/software/asreml.

CERQUERIA, M. D., WEISSMAN, N. J., DILSIZIAN, V., JACOBS, A. K., KAUL, S., LASKEY, W. K., PENNELL, D. J., RUMBERGER, J. A., RYAN, T. and VERANI, M. S. (2002). Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: A statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* **105** 539–542.

EVERITT, B. S. (1995). In the analysis of repeated measures: A practical review with examples. *Statistician* **44** 113–135.

FITZMAURICE, G. M., LAIRD, N. M. and WARE, J. H. (2004). *Applied Longitudinal Analysis*. Wiley, Hoboken, NJ. MR2063401

FRISON, L. and POCOCK, S. J. (1992). Repeated measures in clinical trials: Analysis using mean summary statistics and its implications for design. *Stat. Med.* **11** 1685–1704.

GEORGE, B. (2014). A spatiotemporal model for repeated imaging data. Ph.D. dissertation, Univ. Alabama at Birmingham, Birmingham, AL.

GEORGE, B. and ABAN, I. (2015). Selecting a separable parametric spatiotemporal covariance structure for longitudinal imaging data. *Stat. Med.* **34** 145–161. MR3286245

GEORGE, B., DENNEY, JR., T., GUPTA, H., DELL'ITALIA, L. and ABAN, I. (2016). Supplement to "Applying a spatiotemporal model for longitudinal cardiac imaging data." DOI:10.1214/16-AOAS911SUPP.

GIBBONS, R., HEDECKER, D., WATERNAUX, C. and DAVIS, J. M. (1988). Random regression models: A comprehensive approach to the analysis of longitudinal psychiatric data. *Psychopharmacol. Bull.* **24** 438–443.

KENWARD, M. G. and ROGER, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53** 983–997.

LITTELL, R. C., PENDERGAST, J. and NATARAJAN, R. (2000). Modelling covariance structure in the analysis of repeated measures data. *Stat. Med.* **19** 1793–1819.

MANN, D. L., KEND, R. L., PARSONS, B. and COOPER, G. (1992). Adrenergic effects on the biology of the adult mammalian cardiocyte. *Circulation* **85** 790–804.

MATTHEWS, J. N. S. (1990). A refinement to the analysis of serial data using summary measures. *Stat. Med.* **12** 27–37.

NAGATSU, M., ZILE, M. R., TSUTSUI, H., SCHMID, P. S., DEFREYTE, D., COOPER, G. and CARABELLO, B. A. (1994). Native β-adrenergic support for left ventricular dysfunction in experimental mitral regurgitation normalizes indexes of pump and contractile function. *Circulation* **89** 818–826.

SAS INSTITUTE INC. (2014). SAS/STAT 13.2 User's Guide. SAS Institute Inc.: Cary, NC, 2014.

SCHAALJE, B., MCBRIDE, J. B. and FELLINGHAM, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *J. Agric. Biol. Environ. Stat.* **7** 512–524.

SCHIROS, C. G., DELL'ITALIA, L. J., GLADDEN, J. D., CLARK, D., ABAN, I., GUPTA, H., LLOYS, S. G., MCGIFFIN, D. C., PERRY, G., DENNEY, T. S. and AHMED, M. I. (2012). Magnetic resonance imaging with 3-dimensional analysis of left ventricular remodeling in isolated mitral regurgitation: Implications beyond dimensions. *Circulation* **125** 2334–2342.

SEALS, S., KATHOLI, C. R., ZHANG, J. and ABAN, I. B. (2014). Evaluating the use of spatial covariance structures in the analysis of cardiovascular imaging data. *Comm. Statist. Simulation Comput.* DOI:10.1080/03610918.2014.925922

SIMPSON, S. L., EDWARDS, L. J., MULLER, K. E. and STYNER, M. A. (2014a). Kronecker product exponent AR(1) correlation structures for multivariate repeated measures. *PLoS ONE* **9** e88864.

SIMPSON, S. L., EDWARDS, L. J., STYNER, M. A. and MULLER, K. E. (2014b). Separability tests for high-dimensional, low-sample size multivariate repeated measures data. *J. Appl. Stat.* **41** 2450–2461.

STANDAERT, D. (2014). Image analysis in clinical practice: Measuring progress. Presented at: Advanced imaging analysis. In *Proceedings of the AMC* 21 *Research Strategic Planning Committee Imaging Symposium*; 2014 *Apr* 1. Birmingham, AL.

TSUTSUI, H., SPINALE, F. G., NAGATSU, M., SCHMID, P. S., ISHIHARA, K., DEFREYTE, G., COOPER, G. and CARABELLO, B. A. (1994). Effects of chronic β-adrenergic blockade on the

left ventricular and cardiocyte abnormalities of chronic canine mitral regurgitation. *J. Clin. Invest.*
93 2639–2648.

VALLEJO, G., FERNÁNDEZ, M. P., LIVACIC-ROJAS, P. E. and TUERO-HERRERO, E. (2011). Se-
lecting the best unbalanced repeated measures model. *Behav. Res. Methods* **43** 18–36.

WALLER, L. A. and GOTWAY, C. A. (2004). *Applied Spatial Statistics for Public Health Data.*
Wiley, Hoboken, NJ. MR2075123

YOUNG, A. A., ORR, R., SMAILL, B. H. and DELL'ITALIA, L. J. (1996). Three-dimensional
changes in left and right ventricular geometry in chronic mitral regurgitation. *Am. J. Physiol.
Heart Circ. Physiol.* **271** H2689–H2700.

B. GEORGE   T. DENNEY
I. ABAN   DEPARTMENT OF ELECTRICAL
DEPARTMENT OF BIOSTATISTICS   AND COMPUTER ENGINEERING
UNIVERSITY OF ALABAMA AT BIRMINGHAM   AUBURN UNIVERSITY
1720 2ND AVE S   BROUN HALL, 341 WAR EAGLE WAY
BIRMINGHAM, ALABAMA 35233   AUBURN, ALABAMA 36849
USA   USA
E-MAIL: brgeorge@uab.edu
        caban@uab.edu

                        H. GUPTA
                        L. DELL'ITALIA
                        DEPARTMENT OF MEDICINE
                        UNIVERSITY OF ALABAMA AT BIRMINGHAM
                        1808 7TH AVE S
                        BIRMINGHAM, ALABAMA 35294-0012
                        USA
                        AND
                        BIRMINGHAM VETERAN AFFAIRS MEDICAL CENTER
                        700 SOUTH 19TH STREET
                        BIRMINGHAM, ALABAMA 35233
                        USA