Published in final edited form as:

J Consult Clin Psychol. 2014 October ; 82(5): 760–772. doi:10.1037/a0035267.

Modeling Nonlinear Time-Dependent Treatment Effects: An Application of the Generalized Time-Varying Effect Model (TVEM)

Mariya P. Shiyko,

Department of Counseling & Applied Educational Psychology Bouvé College of Health Sciences,
Northeastern University 404 INV, 360 Huntington Ave Boston, MA 02115 m.shiyko@neu.edu
617-373-7527

Jack Burkhalter, Ph.D.,

Department of Psychiatry & Behavioral Sciences Memorial Sloan-Kettering Cancer Institute

Runze Li, and

Department of Statistics and The Methodology Center Pennsylvania State University

Bernard J. Park, M.D.

Chief of Thoracic Surgery Hackensack University Medical Center

Abstract

Objective—The goal of this paper is to introduce to social and behavioral scientists the generalized time-varying effect model (TVEM), a semi-parametric approach for investigating time-varying effects of a treatment. The method is best suited for data collected intensively over time, e.g., experience sampling or ecological momentary assessment, and addresses questions pertaining to effects of treatment changing dynamically with time. Thus, of interest is the description of timing, magnitude, and (non-linear) pattern of the effect.

Method—Our presentation focuses on practical aspects of the model. A step-by step demonstration is presented in the context of an empirical study designed to evaluate effects of surgical treatment on quality of life among early stage lung cancer patients during post-hospitalization recovery ($N = 59$, 61% female, Mean age = 66.1). Frequency and level of distress associated with physical symptoms were assessed twice daily over a two-week period, providing a total of 1,544 momentary assessments.

Results—Traditional analyses (ANCOVA, repeated-measures ANCOVA, and multilevel modeling) yielded findings of no group differences. In contrast, generalized TVEM identified a pattern of the effect that varied in time and magnitude. Group differences manifested after day four.

Conclusions—Generalized TVEM is a flexible statistical approach that offers insight into the complexity of treatment effects and allows modeling of non-normal outcomes. The practical demonstration, shared syntax, and availability of a free set of macros aim to encourage researchers to apply TVEM to complex data and stimulate important scientific discoveries.

Keywords

time-varying effect; treatment effect; intensive longitudinal data; ecological momentary assessment

Longitudinal research designs are indispensable for examining short- and long-term effects of treatments. Traditional longitudinal studies typically evaluate the overall treatment effectiveness from pre- to post- assessments or monitor treatment progress based on weekly or monthly retrospective recalls. The process of change, however, is rarely instantaneous. Time-intensive data with frequent assessments recorded in real-time and real-life settings provide novel opportunities to understand the mechanism and patterns of change, duration of the treatment effect, and periods of fallbacks and recovery. For example, effects of a cognitive-behavioral therapy (CBT) for patients with an eating disorder can be effectively assessed through electronic diaries capturing binge-eating events and emotional stability in real time prior to and following a treatment (Munsch et al., 2009). Similarly, changes in the functioning of children with an attention-deficit/hyperactivity disorder (ADHD) due to a pharmacotherapy are evident from assessments taken in a real-life environment full of demands and stressors (Whalen et al., 2010). Such studies capture nuances and challenges of daily functioning in real or near real time and reflect realities of treatment effects in the context of daily living.

With high data resolution and rich within-person records, there is a great need for statistical methods that accommodate the analytical challenges of such data, including their large volume, incomplete data, unbalanced times of measurement, different measurement scales, and complex (usually, non-linear) patterns of change. The goal of this paper is to introduce the generalized time-varying effect model (TVEM; introduction to TVEM for normally distributed outcomes can be found in Shiyko, Lanza, Tan, Li, & Shiffman, 2012; Tan, Shiyko, Li, Li, & Dierker, 2012) that uniquely addresses challenges of intensive longitudinal data (ILD) and enables examination of novel research questions about micro processes of change in the context of a treatment. The empirical demonstration, a step-by-step presentation of the model, and a reference to a free set of macros are intended to stimulate applications of the method by applied researchers.

Incentives for Assessing Treatment Effects Intensively and in Real Life

There are several notable advantages of using daily and momentary methods of data collection to capture treatment effects. In the area of clinical psychology, outcomes of interest such as depression, anxiety, cigarette and food cravings, medication adherence, quality of social interactions, quality of life, negative and positive affect, are usually measured by self-report. However, global ratings of behaviors, psychological phenomena, and context have been found to suffer from cognitive biases such as event recency and saliency, personal affective state, and fatigue, among others (Heron & Smyth, 2010; Shiffman, 2009; Sohl & Friedberg, 2008; Stone, Schwartz, Broderick, & Shiffman, 2005). In comparison, momentary ratings with a temporal focus on the present moment, e.g., “How anxious are you right now?” tend to reduce the recall bias and, through repeated sampling, capture temporal sequencing of a phenomenon as it unfolds (Shiffman, 2005; Shiffman,

Stone, & Hufford, 2008). Further, global self-assessments may not emulate real-life experiences or provide accurate reports of whether a treatment effected actual behavior change. Research has demonstrated that effectiveness of a treatment can be overlooked with global measures (Barge-Schaapveld & Nicolson, 2002; Gehricke, Hong, Whalen, Steinhoff, & Wigal, 2009; Volkers et al., 2002) or its effect is detected with a delay (Lenderking et al., 2008). Assessments taken in vivo (e.g., with smart phones, heart-rate monitors), on contrary, are more sensitive to subtle improvements (Ebner-Priemer & Trull, 2009; Myin-Germeys et al., 2009; Walls & Schafer, 2006) and help to assure ecological validity of an observed treatment effect. Finally, from the standpoint of study design, frequent assessments offer an opportunity to capture the mechanism of change and timing of a treatment effect.

Recognizing multiple benefits of real-time assessments, the U.S. Food and Drug Administration (U.S. Food and Drug Administration, 2009) issued a recommendation for promoting this method in documenting treatment progress and outcomes. In addition, the method of real-time data capture was put forward as a priority for development by the U.S. Department of Health and Human Services due to its “proactive, vigilant, and risk reducing” potential (p.20, Honoré, & Scott, 2010) to prevent and improve interventions for mental and physical illness. Because of this enthusiasm, numerous advantages, and continuous technological innovations, a number of studies that collect ILD are on the rise, placing a higher demand on statistical tools for flexible and informative unpacking of rich data.

Modeling Time-Varying Treatment Effects

Recognizing the importance of longitudinal studies and addressing unique demands of repeated assessments, a number of volumes have been published on methods for modeling change (e.g., Bollen & Curran, 2006; Collins & Sayer, 2001; Hedeker & Gibbons, 2006; Nagin, 2005; Preacher, Wichman, MacCallum, & Briggs, 2008; Singer & Willett, 2003). However, ILD in social and behavioral sciences are unique with the respect to the large number of assessments (frequently, more than 20 or even 100 per person), person-specific number and timing of measurement points, incomplete data, and complex (not-linear) temporal fluctuations. These characteristics preclude a direct application of traditional longitudinal methods to ILD. Investment in specialized analyses and exploration of ILD richness, however, opens possibilities for exploring novel questions about nuances of within-person changes emerging over time. In the context of a treatment, ILD can help to uncover and describe a response pattern as it unfolds in time.

The simplest approach to ILD analysis is data aggregation (e.g., computing an average response based on momentary records) to fit requirements of traditional statistical methods (e.g., t-test, pre- and post- test ANOVA). While important for initial exploration, such aggregations frequently mask temporal fluctuations of the response and address only global questions of change. A more sensitive approach is to preserve the order and frequency of original assessments and apply techniques for modeling the pattern of response. Latent growth curve (LGC) modeling is widely used for studying change in traditional panel studies. It accommodates a variety of response patterns, including variations of linear (e.g., linear, quadratic, cubic, and piece-wise; Bollen & Curran, 2006; Browne & du Toit, 1991; McArdle, 2009; Preacher, Wichman, MacCallum, & Briggs, 2008) and non-linear forms

(e.g., latent basis LGC; McArdle & Epstein, 1987; Meredith & Tisak, 1990; structured latent curves; Brown, 1993). Multi-group specification of the model (McArdle, 1989; McArdle & Bell, 2000, McArdle & Epstein, 1987) makes it possible to compare features of trajectories between groups, such as the level of response at a given time point or the rate of change. For freely-varying trajectories, as in latent basis LGC, group comparisons are not intuitive and not easily interpretable. In its classical form, LGC engendered few applications with ILD. Due to recent developments, some extensions are available for time-series data (still slightly different from ILD) to address questions about the autoregressive pattern (Molenaar, Sinclair, Rovine, Ram, & Corneal, 2009).

Multi-level modeling (MLM) has become a popular alternative to LGC and is the most widely used method for ILD analyses (Schwartz & Stone, 1998, 2007). MLM parallels some specifications of LGC (Hox & Stoel, 2005; Kaplan, 2009). MLM is particularly suitable for ILD because of its flexibility to handle data features, research focus on estimating population- (fixed) and person- (random) level parameters, inclusion of time-varying and time-invariant covariates, and extensions to non-normally distributed outcomes. In generalized MLM, between-group comparisons in outcome trajectories can be specified as an interaction between group and time variables: $Group_i * Time_{ij}$. If time is measured continuously, the effect of treatment is tested along the entire temporal scale. Time, then, serves as an effect modifier that captures a pattern of increasing improvement due to treatment or a progressive diminishment of the effect (e.g., Wing et al., 2008). The caveat, however, is that time is expressed parametrically and, frequently, linearly, thus imposing a big assumption of a constant rate of change between the groups. Other more complex parameterizations can include piece-wise linear (Husson et al., 2008), quadratic, logistic, and exponential (see Bellera et al., 2010, in a context of survival analysis). These extensions add flexibility to the shape of the model at the expense of increased challenge of interpretation. As other parametric methods, MLM is sensitive to model misspecification.

In contrast to traditional modeling techniques, generalized TVEM introduced in this paper is a semi-parametric model that requires no predetermined specification on the shape of change. Thus, response over time can follow a complex pattern, and there is no need to define a priori moments where treatment might begin to manifest or lose its potency. Instead, group differences over time are directly observed, estimated, and compared from the data. The work on TVEM is grounded in the theory of varying-coefficient models (Hastie & Tibshirani, 1993; Hoover et al., 1998), but only recently has the method become available to applied scientists. This recent development also extended the method to applications with ILD, allowing for incomplete and unbalanced observations (Shiyko, Lanza, et al., 2012, Tan et al., 2012, Yang et al., 2012). In the following sections, to assure a user-friendly presentation of the method, we begin with an empirical example of data from a study of lung-cancer patients recovering after hospitalization from either of two types of surgical treatment. Within this empirical demonstration we focus on the: 1) theoretical formulation of research questions; 2) data preparation; 3) model specification; 4) model fitting; and 5) result summary and interpretation. Practical recommendations follow in the discussion section.

Empirical Example: Post-Surgical Recovery in Lung Cancer Patients

Our empirical example comes from a behavioral science correlative study that examines quality of life in surgically treated lung cancer patients. Fifty-nine patients diagnosed with early stage lung cancer (61% female, 73% married, Mean age = 66.1 years old, SD = 7.9) underwent one of two types of surgical procedures: a traditional open thoracotomy and lobectomy (THOR, 39% of patients) that utilizes a single, large, rib-spreading incision for direct visualization, or a minimally invasive video-assisted thoracic surgery and lobectomy (VATS; 61%) that requires three small, non-rib spreading incisions and videoscopic guidance to accomplish the same operation. Because of the nature of the treatment, random group assignment was not feasible. Groups did not differ on any generally recognized confounders, and they were dropped from subsequent analyses. The goal of the study was to examine the relationship between type of surgery (VATS vs. THOR) and physical symptoms and their impact on daily functioning and life quality after discharge from the hospital. Given the relatively good survival prognosis for surgically treated, early-stage lung-cancer patients, assessing aspects of quality of life, such as symptom burden and physical function, is important in informing efforts for optimizing care of these cancer survivors (Kenny, et al., 2008; Li, Lee, & Yim, 2004).

A modified Memorial Symptom Assessment Scale (MSAS, Portenoy et al., 1994) was used to assess the frequency and burden of physical symptoms common for post-surgical lung-cancer patients twice daily (morning and afternoon) with electronic portable devices over the course of two weeks. The temporal focus of questions was on the present moment and included symptoms of pain, e.g., “Right now, do you have pain?” (yes/no) and, if yes, “Right now, how much does this pain bother you?” (not at all – 0 – to very much – 4), energy level, drowsiness, appetite intensity, constipation, shortness of breath, and chest tightness. To accommodate the highly skewed nature of the scores, two scales were constructed: the frequency of symptoms scale (MSAS_F, ranging 0 to 8), indicating the total number of symptoms present, and the symptom burden scale (MSAS_B, ranging 0 to 8) constructed by counting the number of symptoms that were endorsed as burdensome, from a little to very much. For 81.4% of study participants, adherence to momentary responses was perfect, yielding 28 assessments per person (2 times a day x 14 days). The remaining participants provided 8 to 26 data points (Mean = 18.2, Median = 20, SD = 6.1). Missing data occurred due to early study withdrawals, with 8 participants in the THOR and 3 in the VATS groups ($\chi^2 = .29, p = .59$) withdrawing, on average, on 10th day (Median = 11, SD = 3.05). We controlled for the early drop-out status in subsequent analyses.

In our initial analyses, we conducted group comparisons based on three of the most frequently used approaches: a comparison of group means through ANCOVA or repeated-measures ANCOVA, and a comparison of group trajectories through MLM. To compare group means, all momentary responses were averaged, yielding one record per individual. Based on the ANCOVA results, no group differences were found either in symptom burden ($\overline{MSAS_B}_{THOR} = 2.54, SD = 2.05; \overline{MSAS_B}_{VATS} = 1.97, SD = 1.71; F_{1,56} = .311; p = .579$) or symptom frequency ($\overline{MSAS_F}_{THOR} = 2.87, SD = 1.65; \overline{MSAS_F}_{VATS} = 2.66, SD = 1.62; F_{1,56} = .241; p = .626$). Participants with earlier drop-out were equivalent on

symptom burden to those with full compliance ($p = .121$) but had an elevated symptom count by about 1 point ($p = .068$). For repeated-measures ANCOVA, we computed initial and final MSAS scores based on the first and last three days of assessment, respectively. Three days (6 assessments) were chosen to obtain stable estimates and minimize missing data, although missing data were unavoidable due to attrition towards the end of the study. This choice demonstrates decision-making challenges that go into the process of data reduction. Based on the results, only the effect of time was found to be significant both for symptom frequency ($F_{1,96} = 10.059, p = .002$) and burden ($F_{1,96} = 6.024, p = .015$), corresponding to a reduction about one point (for $\overline{MSAS_F}$ from 3.49 to 2.29, and for $\overline{MSAS_B}$ from 3.03 to 2.01), irrespective of the treatment group.

Further, generalized MLM (with a log link function for a Poisson-distributed outcome) was fitted with two interaction terms ($Group_i * Time_{ij}$ and $Group_i * Time_{ij}^2$) to model possible time varying, between-group differences. No such differences were found as indicated by model fit indices (Model 1 with homogenous slopes: BIC = 956.53, AIC = 922.39; Model 2 with group-specific slopes: BIC = 976.85, AIC = 928.07) and tests of parameter estimates (all $p > .8$ and coefficients for group differences close to 0). Results were consistent irrespective of the parameterization of time (e.g., centered at the first or last assessment points). For both outcomes, MSAS_B and MSAS_F scores were found to progressively decrease over time in a slow quadratic fashion for all individuals ($p < .05$ for linear and quadratic time effects). To summarize, all of the above approaches failed to find any group differences, although repeated-measures ANCOVA and MLM captured reduction in symptoms and burden over time.

These findings are further contrasted with results from TVEM. To capitalize on repeated assessments of the outcome through the application of TVEM we set to answer the following research questions: a) Are there differences between the treatment (VATS) and control (THOR) groups during the two-week period of post-surgical recovery? b) Is the treatment effect time-varying? and c) What is the pattern of the treatment effect?

Overview of TVEM

Model Specification

In this paper, generalized TVEM is presented for the audience of applied researchers interested in examining time-varying treatment effects. A more general introduction can be found in Tan et al. (2012). First, intensively measured outcome can be defined as follows:

$$\{Y_{ij}\}, \quad i=1, \dots, n; \quad j=1, \dots, m_i,$$

where Y_{ij} is the outcome for participant i measured at time t_{ij} . The total number of observations for participant i , denoted by m_i , can be unique for each person. In addition to the intensively measured outcome, each participant has a treatment assignment, Tx_i , and time-invariant covariates, X_{ip} (e.g., gender, age, where p indexes the individual covariates). In a data spreadsheet, ILD are summarized in a long data format such that each observation

is uniquely defined by a subject id and a time stamp (assuming time is measured on a continuous scale).

Further, Generalized TVEM with a single covariate takes the following form:

$$g\{E(Y_{ij})\} = \beta_0(t_{ij}) + \beta_1(t_{ij}) * Tx_i + \beta_2 * X_i,$$

where $g\{E(Y_{ij})\}$ stands for a link function $g(\cdot)$ for non-normally distributed outcomes (e.g., binary, count, zero-inflated, see Yang et al., 2012). The intercept $\beta_0(t_{ij})$ and slope $\beta_1(t_{ij})$ are defining features of TVEM and can be understood in a context of traditional regression. In a regression model, parameters β_0 and β_1 are estimated through ordinary least squares and take constant values. For example, with two treatment groups where Tx_i takes values of 1 for treatment and 0 for control groups, the intercept, β_0 , represents the mean outcome scores for the control group; and the slope, β_1 , captures the mean difference in the outcome between treatment and control groups. When the value of the slope is significantly different from 0, we conclude that a treatment had an effect, or, to be more precise, we find no evidence for a null hypothesis that group means are equivalent. Covariates may be entered in the model as control variables. In comparison, in TVEM, both $\beta_0(t_{ij})$ and $\beta_1(t_{ij})$ are functions defined by t_{ij} that take different values depending on where in time the groups are being compared. While the meaning of the intercept and slope parameters stay the same, the groups now are being compared across the entire time continuum (e.g., two study weeks). Thus, between-group comparisons can be made about the shape, magnitude, and directionality of the effect. Of note, in TVEM effect of covariates can be specified as time-invariant or time-varying (for examples, see Shiyko, Lanza, et al., 2012; and Shiyko, Naab, Shiffman, & Li, 2013).

Model Estimation

TVEM is a semi-parametric model, and its parameter functions can take a variety of continuously changing shapes (e.g., without abrupt break points). The goal, then, is to estimate these functions. We present a conceptual overview of the estimation process and refer to a hypothetical function in Figure 1. Multiple approaches for curve estimation have been proposed and discussed in the statistical literature; interested readers are referred to Tan et al. (2012). Here, we focus on the spline-based method that has been implemented in a freely available set of SAS macros (the macros and an accompanying user guide can be found on <http://methodology.psu.edu>). The function in Figure 1 is complex and cannot be easily estimated with a line or quadratic or cubic curves. However, it can be truncated into smaller intervals by placing dividing points (knots) either equidistantly or at selected points along the continuum. Vertical lines in Figure 1 demonstrate how the curve can be split into seven equal intervals with six knots. Within each smaller interval, a simpler function (e.g., quadratic) can be used to model what is now a simpler and familiar shape of a curve. Thus, the spline method refers to such piece-wise modeling of a complex function with polynomials (linear, quadratic, or cubic) on its smaller intervals. Our simulation study (Tan et al., 2012) demonstrated the accuracy of this approach in estimating curves of varying complexity. The key is to balance between an adequate number of knots for complex functions and over-fitting the model.

The SAS macros implement two spline-based methods: the B- and P- splines. The major difference between the two is the degree of smoothness in resulting curves. The B-spline approach largely follows the logic outlined above, and the process of model selection requires an iterative fitting of models with a different number of knots. For example, one may start with 5 knots and, depending on the complexity of estimated functions, gradually increase or decrease the number. The best-fitting model is, then, selected based on the fit statistics (more details are to follow). Depending on the chosen number of knots, the B-spline method yields rougher functions that match the data more closely. This is an advantage when a more fine-grained examination of the observed data is a priority. With the B-spline approach, it is also possible to compare TVEM with simpler parameterizations of functions (e.g., quadratic, linear), especially when model fit indices are in favor of simpler models (see Shiyko, Lanza, et al., 2012, for an example). The drawback, however, is data over-fitting. In many applications in psychology there is an inherent measurement error, and recovering small fluctuations in parameters may be misleading or unimportant. In summary, the B-spline approach may be useful when the shape of the modeling curves is rather simple (e.g., monotone, without complex ups and downs), when a direct comparison with parametric functions is of interest, and when data quality is not a concern.

In contrast, the P-spline method simplifies the model-fitting process. It imposes a quadratic penalty (P) term to avoid data over-fitting, yielding smoother parameter estimates. Smoothing is carried out by shrinking coefficients of fluctuation at every knot, and its degree is determined by estimates from the mixed-effects model (see Rupert, 2002, and Wand, 2003, for details). The method uses a tuning parameter in the penalty to control the model complexity. Due to the quadratic form of the penalty function, degrees of freedom (i.e., the number of effective parameters) are computed as the trace of projection matrix of the linear smoother corresponding to the P-spline (Rupert, 2002). Based on simulation studies, 5 to 10 knots are sufficient to estimate functions changing in a monotone fashion or functions with a single mode. Multimodal functions with 3 to 6 cycles are better approximated with 20 knots, while very complex functions with more than 10 modes require at least 40 knots (Rupert, 2002). In social sciences, diurnal cycles may interfere with a treatment response and produce highly fluctuating response patterns. From our experience, however, most applications produce relatively simple monotone or unimodal shapes, and 10 knots is generally sufficient to capture their complexity (Tan et al., 2012). Overall, the P-spline estimation method is more straight-forward and requires only the initial decision on the number of knots. Generally speaking, parameter functions estimated by both methods are similar in shape, although the B-spline results in wigglier curves that capture every nuance of fluctuation, while the P-spline smoothes fluctuations avoiding data over-fitting. The two methods can be viewed as complimentary and selected depending on the analytical purpose and data quality.

Independent of the estimation method, TVEM parameter functions represent population-based estimates, similar to fixed effects in MLM or parameter estimates in generalized estimation equations. While somewhat restrictive, this assumption allows ILD to be used effectively. In contrast to functional or time-series data, where participants are assessed on the same schedule, EMA studies frequently yield incomplete and unbalanced data, and knowing responses from all participants in the same group allows for more accurate

estimates of the curves. This has direct design implications, where a good balance is needed between the number of participants (with implications for estimating curves accurately with a smaller degree of uncertainty), the number of assessments (generally speaking, the more the better, especially when the process of change is highly dynamic), and timing of assessments (different times across participants are of benefit for capturing nuances of the curve at different intervals). In discussion, we return to the issue of design in more detail. Conceptualizing parameter functions as population-level effects also simplifies the problem of knot selection. Although the model is developed for ILD, individuals who contribute only a few assessments (e.g., < 10) are still of value, and do not impact the decision for the number of knots. Of note, the current version of SAS macro assumes working independence correlation structure for the error process, but the standard errors are estimated by the robust sandwich formula to implicitly take into account the error correlation structure (Huber, 1967; White, 1982).

Finally, besides the estimates of the effect curves, TVEM provides 95% point-wise confidence intervals (CI) for each effect curve. These intervals may be used for point-by-point comparisons, e.g., do groups differ on day 3? However, they may be inappropriate for examining overall group differences, e.g., do group differ over the entire time interval? To avoid inflation of type-I error, it is advisable to establish comparison points *a priori* or carry out analysis in an exploratory manner, with the goal of subsequent replications. While preferred, simultaneous CIs are difficult to construct and are subject to future research.

Model Fitting

In summary, the following steps should be taken for fitting Generalized TVEM to data:

1. Investigate the nature of the outcome distribution (e.g., count, binary) in order to select an appropriate link function in Generalized TVEM.
2. Summarize data graphically through group profiles and examine complexity of curves.
- 3a Use the P-spline Generalized TVEM, setting the number of knots to 10 for simpler curves and 20 or more for more complex curves. No model selection is required.
- 3b Alternatively, or as a complement, use the B-spline Generalized TVEM. Set the number of knots to 5 and progressively increase or decrease the number. Use fit indices, e.g., lowest values on AIC, Corrected AIC, and BIC, to select the best-fitting model.
- 4 Summarize results graphically and interpret.

Application of Generalized TVEM to Empirical Data

In this section we demonstrate an application of Generalized TVEM to data on postoperative recovery of lung cancer patients who underwent either of two types of surgical treatments. Figures 2 and 3 summarize, respectively, symptom frequency and symptom burden trajectories for 10 randomly selected individuals from the two surgical treatment groups. The solid thick lines reflect mean trajectories for each group. It is apparent that recovery

scores follow a non-linear pattern over the two weeks of observation. Patients in the VATS group also appear to have somewhat lower symptom frequency and symptom burden scores, with clearer group differences manifesting within the first week of the study.

To investigate group differences and their time-varying nature, we fit Generalized TVEM (with a log link function to account for the Poisson-distributed outcomes):

$$\log (E (MSAS_{ij})) = \beta_0 (t_{ij}) + \beta_1 (t_{ij}) * Tx_i + \beta_2 * X_i.$$

In the dataset, group membership was coded such that $Tx_i = 0$ for THOR and $Tx_i = 1$ for VATS patients. Comorbidities (present = 1 for 59.3% of patients, absent = 0) and age (centered at Median = 65 years) were initially included as covariates; however, these covariates were not found to be significant and were omitted from the final models.

As part of the demonstration, we present results from both for B- and P-spline estimation methods for the outcome of symptom frequency. When fitting the model with the B-spline approach, the initial number of knots was set to 5 (see Appendix 1 for SAS syntax). Then, the model was progressively simplified, and fit indices were recorded (Table 1). Ideally, all fit indices should agree on the final solution, but frequently they point towards different results, especially in cases of small sample data. For example, BIC (Schwarz, 1978) is designed to select simpler models and penalize models with a larger number of parameters, e.g., models with a higher number of knots. AIC and Corrected AIC (AICC; Sugiura, 1978), in comparison, may prefer more complex models. Further, simulations show that AICC is preferential for small-sample data and nonlinear regression applications (Hurvich and Tsai, 1989). Thus, relying more heavily on AIC and AICC in our analyses, the final model had 2 knots for both intercept and slope functions. In comparison, model fitting with the P-spline method was simpler and entailed only setting the number of knots to 10. For symptom burden, we present results only for the P-spline estimation method.

The intercept functions (Panels A in Figures 4 - 6) capture trajectories of the level of burden and number of symptoms over time for the THOR group. At the first day of observation, THOR patients were estimated to report about 4 symptoms (Panel A, Figures 4 & 5). This number decreased by over a point after four days in the study, plateaued, and further decreased on day 10. The results were consistent irrespective of the estimation method, although the P-spline approach yielded smoother curves, as expected. A slight decrease in symptom burden (Panel A, Figure 6) was also evident, with scores reducing from 3 to 2.5 over the course of two weeks, with the reduction occurring only in the first week of observation.

The slope functions (Panels B in Figures 4 - 6) capture the effect of VATS on symptom frequency and symptom burden relative to THOR. Because of the parameterization of the model to fit the Poisson outcome, the slopes are scaled to risk ratio and are factors by which VATS group differs from THOR at each time point. Once significantly different from and less than one, the effect of VATS manifests in a reduced number of symptoms and burden. The confidence bands (thinner curves) around the estimated averages (thick middle curve) inform whether and where the groups differ from each other. Based on the summaries, group

differences in the number of symptoms are not present until the end of day four, after which we observe a progressively larger improvement in the VATS group (Panels B in Figures 4 & 5). While curves converge at the end, this is likely an artifact of the boundary effect, common in spline-based estimation methods (Härdle, 1990). In addition, on day 3, patients in the VATS group begin to experience a lighter burden from physical symptoms (Panel B, Figure 6), with progressively growing group differences over the course of the first week and maintenance during the second week. To aid interpretation further, Panels C summarize estimated mean curves for both groups. From these graphs, it is apparent that the effects of the two types of surgery are not homogenous and vary in time and magnitude.

Discussion

Increased enthusiasm and knowledge about real-time and real-life methods of data collection, the availability of improved and inexpensive technology, and ease of data storage have been catalysts for testing the impact of clinical treatments in situ. Continuous and intensive data collection offers fresh glimpses into the unknowns of behavior change processes – not from the perspective of a retrospective recall or theoretical speculation – but from a direct capture of personal experiences based on frequent real-time assessments. Such methodology is of great importance for examining treatment effectiveness and its temporal dynamics and nuances.

The goal of the current paper was to introduce the generalized time-varying effect model (TVEM) to behavioral and social scientists in order to offer a specialized and flexible tool for assessing treatment effects recorded intensively over time. Our presentation focused on the conceptual and practical overview of the model to stimulate further applications among those collecting ILD. It is instructive to highlight the unique features of TVEM. First, it is a semi-parametric method that offers flexibility in modeling complex temporal shapes. Second, TVEM is robust to the issue of model misspecification, which can be a concern for parametrically specified coefficient functions (Hastie & Tibshirani, 1993). Third, the method is well suited to handle nuances of ILD. Finally, most real-time data collection designs follow a regimen of random prompts or participant-initiated responses (Shiffman, 2007), making unfeasible an alignment of observations across participants that is required for traditional statistical methods. In contrast, TVEM benefits from such temporal dispersion of data points, since functions represent population-level estimates and draw from all data. While temporal staggering of data due to design is not a problem, a systematic bias due to data missingness can skew the results. Thus, researchers are advised to examine patterns of missingness to be informed about inherent data biases and related limitations in inferences. Our empirical example was based on a fairly complete set of assessments; however, future research should examine effects of data missingness on TVEM estimates under different conditions.

The issue of data quality deserves a special discussion. Given that no constraints on the shape of TVEM parameter functions are imposed, estimation is largely data driven. While this is not unique to TVEM, its semi-parametric nature requires data of high quality to ensure valid and generalizable results. Overall, principles of good research practices apply to studies of real-time and real-life experiences, but there are a number of special considerations (A.

Stone & Shiffman, 2002). They include unique requirements of the questionnaire development, its repetitive nature, limited time of administration, and unique validation practices (Shrout & Lane, 2012). The concept of treatment fidelity is also extended to include an evaluation of technology and provision of technical assistance (DeVito Dabbs et al., 2011). Data quality is also dependent on the chosen sampling design, which should be carefully considered and reflect the nature of the phenomena under investigation, their dynamic properties, rates and patterns of change in such phenomena, and the issue of subject burden (Collins, 2006; Shiyko & Ram, 2011). Our empirical example was carried out with 59 participants and a total of 1,544 momentary assessments, which demonstrates feasibility of applications with relatively small samples. Our general recommendation is to have a good balance between the number of assessments for each participant and the number of participants. Both influence the accuracy of parameter estimates and, therefore, have impact on CIs as well as power of hypothesis testing. In general, the number of participants will have more direct influences on estimation accuracy, since participants are independent from each other, while observations are not. Further, rates and patterns of adherence to assessment schedules are of great importance (DeVito Dabbs et al., 2011). Finally, a strong theoretical foundation should accompany decisions about study design and data analyses.

We want to emphasize the nature of research questions suitable for TVEM. Our empirical demonstration focused on the sequelae of two types of surgical procedures with lung cancer patients' post-operative recovery, as indicated by the frequency and burden of physical symptoms. We compared effects of the type of surgery during the period of two weeks following hospital discharge. We established a time point where treatment effects began to manifest (about 3 or 4 days into the study) and observed the extent of diverging patterns of treatment effects. When a specific time point is of particular interest for clinical purposes, one can clearly infer about group differences at that point (or region); however, the number of such points should be kept small to avoid the issue of large-scale multiple comparison. Further, critical periods can be identified from TVEM results and targeted by adding treatment and prevention components. General patterns of group differences are also of great importance. Frequently, effects of a treatment may take time to manifest, and TVEM analysis can help to identify the period of latency. In addition, changes in the strength of the effect can be explicitly examined, with direct implications for clinical practices. For example, in this study, the data suggest that VATS was associated with lower symptom experience and burden that began to manifest a few days after hospital discharge and were mostly apparent after one week. These findings are of importance to future patients who can be reassured about the recovery process and assisted more intensively in the first few days following hospital discharge. Of great importance is the fact that group differences were overlooked with traditional methods of modeling ILD, i.e., ANCOVA, repeated-measures ANCOVA, and the multilevel approach. In other applications, such methods may pick up an effect, but produce biased estimates, depending on the shape of the parameter functions.

In the current application of TVEM, the focus was on the time-varying effects of treatment. Additional extensions are possible, where both a predictor (e.g., pain level) and an outcome (e.g., quality of life) are assessed on the ILD scale, and the association between two or more intensively sampled variables is modeled as a function of time (or some other meaningful effect modifier; Shiyko, Lanza, et al., 2012; Shiyko et al., 2013). The simultaneous

estimation of group trajectories and the focus on the time-varying effects that follow a complex non-linear pattern is a unique feature of TVEM that sets the model apart from other semi-parametric approaches such as, for example, generalized additive models (Hastie & Tibshirani, 1986).

With TVEM, comparisons can be made between several treatment groups, covariates can be explicitly added to the model, and inter-individual differences can be explored either through specification of group membership or covariate influences. A more subtle heterogeneity can also be examined through the inclusion of random effects (Li, Root, & Shiffman, 2006), but this model specification is currently not available in the current SAS macros. Inclusion of random effects would pose additional requirements on the sample size and data properties that are not always available in EMA designs. While the focus of this paper is on ILD collected through studies that employ methods such as ecological momentary assessments or experience sampling, ILD may also arise in other applications. For example, longitudinal studies that span over years and capture phenomena with sufficient frequency, e.g., life-span studies, can also fruitfully apply TVEM.

To ensure a valid model implementation, researchers should pay a close attention to the measurement scale of the modeled outcome (see a discussion in Shiyko, Li, & Rindskopf, 2012). Current modeling capacities of TVEM include outcomes measured on the continuous normal, binary, Poisson, and inflated Poisson scales, thus allowing much flexibility for modeling diverse phenomena. Our empirical demonstration sets an example of modeling count data, which are frequently used in studies of addictive behaviors, e.g., number of smoked cigarettes or consumed alcoholic drinks; measures of symptomatology, e.g., number of binge or purge episodes, frequency of self-harming thoughts, and; behavior frequency, e.g., number of employed coping strategies.

Finally, TVEM can be applied in conjunction with other methods. As an exploratory tool, it can aid in understanding complex treatment responses and suggest potential parametric alternatives if they are preferred. TVEM can also be used as a stand-alone method that summarizes dynamic response patterns in an intuitive way, facilitating direct group comparisons. Since TVEM and Generalized TVEM have not been extensively applied to ILD (some exceptions include Shiyko, Lanza, et al., 2012; Shiyko et al., 2013), more methodological work is needed to understand issues related to power and measures of effect size.

Concluding Remarks

Many theories of mental and physical health presume a dynamic nature of human functioning, where changes occur in time and in natural settings (e.g., Ebner-Priemer & Trull, 2012; Glanz, Rimer, & Viswanath, 2008). Assessment of treatment effects in situ offers an exciting opportunity for ecologically valid inferences and deeper understanding of treatment progression. Our work summarized in this paper is a response to increasing demands for statistical methods that allow nuanced exploration of treatment effects and possess flexibility for handling the challenges of ILD. We encourage applied researchers to use the method of TVEM with the purpose of promoting scientific development and ensuring meaningful and efficient analyses of treatment effects.

Acknowledgments

This study was supported by the National Cancer Institute grant 1R03CA171809-01 and the National Institute on Drug Abuse grants P50 DA010075 and R21 DA024260, T.J. Martell Foundation; the Cancer and Leukemia Group B (now the Alliance for Clinical Trials in Oncology) PHS U10 CA31946; the Memorial Sloan-Kettering Cancer Center Thoracic Service Steps for Breath Fund, and the National Cancer Institute T32CA009461. We gratefully acknowledge Dr. Yuelin Li for comments on an early version of this manuscript and Christopher S. Webster for administrative support.

Appendix

Appendix 1

Summary of SAS code for the empirical example: Time-varying treatment effects of two types of surgery on symptom frequency and burden in recently treated lung cancer patients.

Step 1.

Download SAS macro for Generalized TVEM from <http://methodology.psu.edu>

Step 2.

Read in a data set in SAS.

Step 3.

Create a vector of 1s for intercept function.

DATA example;

SET example;

x0 = 1;

RUN;

Step 4 a.

Define the model and run TVEM with P-spline.

```
%TVEM_poisson (method = P-spline, mydata =
example, id = SubjID, time = TimeSinceSx, dep =
MSAS, tcov = x0 VATS, cov = age, cov_knots = 10
10);
```

```
Call the macro
Method of model estimation
Specify data set
Specify ID variable
Specify time indicator
Specify count dependent variable
Specify names of time-varying parameters (intercept, group)
Specify names of time-stable covariates (age)
Number of knots for all time-varying (tcov) parameters:
intercept (x0) and slope (VATS) functions
```

Step 4 b.

Alternatively, define the model and run TVEM with B-spline.

```
%TVEM_poisson (method = B-spline, mydata =
example, id = SubjID, time = TimeSinceSx, dep =
MSAS, tcov = x0 VATS, cov = age, cov_knots = 2
2);
```

```
Call the macro
Method of model estimation
Specify data set
Specify ID variable
Specify time indicator
Specify count dependent variable
Specify names of time-varying parameters (intercept, group)
Specify names of time-stable covariates (age)
Number of knots for all time-varying (tcov) parameters:
intercept (x0) and slope (VATS) functions
```

References

- Bachrach, CA. [September 22, 2012] Understanding health and behavior in bits and bytes. 2010. from obssr.od.nih.gov/about_obssr/from_the_director/previous_columns.aspx
- Barge-Schaapveld D, Nicolson N. Effects of antidepressant treatment on the quality of daily life: An experience sampling study. 2002; 63(6):477–485.

- Bellera CA, MacGrogan G, Debled M, Tunon de Lara C, Brouste V, Mathoulin-Pelissier S. Variables with time-varying effects and the cox model: Some statistical concepts illustrated with a prognostic factor study in breast cancer. *BMC Med Res Methodol*. 2010; 10(20)
- Bollen, KA.; Curran, PJ. *Latent Curve Models: A Structural Equation Perspective*. John Wiley & Sons, Inc; Hoboken, NJ: 2005.
- Browne, MW. Structured latent curve models.. In: Cuadras, CM.; Rao, CR., editors. *Multivariate analysis: Future directions 2*. North-Holland; Amsterdam: 1993. p. 171-198.
- Browne MW, Du Toit SHC. Automated fitting of nonstandard models. *Multivariate Behavioral Research*. 1992; 27:269–300.
- Collins L. Analysis of longitudinal data: The integration of theoretical model, temporal design, and statistical model. *Annual Review of Psychology*. 2006; 57:505.
- Collins, LM.; Sayer, AG. *New methods for the analysis of change. decade of behavior*. American Psychological Association; Washington, DC, US: 2001.
- DeVito Dabbs A, Song M, Hawkins R, Aubrecht J, Kovach K, Terhorst L, Callan J. An intervention fidelity framework for technology-based behavioral interventions. *Nursing Research*. 2011; 60(5): 340–347. 10.1097/NNR.0b013e31822cc87d. [PubMed: 21878796]
- Ebner-Priemer UW, Trull TJ. Ecological momentary assessment of mood disorders and mood dysregulation. *Psychological Assessment*. 2009; 21(4):463. [PubMed: 19947781]
- Ebner-Priemer, UW.; Trull, TJ. Investigating temporal instability in psychological variables.. In: Mehl, MR.; Conner, TS., editors. *Handbook of research methods for studying daily life*. Guilford Press; New York, NY: 2012. p. 423
- Gehricke J, Hong N, Whalen CK, Steinhoff K, Wigal TL. Effects of transdermal nicotine on symptoms, moods, and cardiovascular activity in the everyday lives of smokers and nonsmokers with attention-deficit/hyperactivity disorder. *Psychology of Addictive Behaviors*. 2009; 23(4):644. [PubMed: 20025370]
- Glanz, K.; Rimer, BK.; Viswanath, K. *Health behavior and health education: Theory, research, practice*. John Wiley & Sons, Inc.; 2008.
- Härdle, W. *Applied non-parametric regression*. Cambridge University Press; Cambridge, UK: 1990.
- Hastie T, Tibshirani R. Generalized additive models. *Statistical Sciences*. 1986; 1(3):297–318.
- Hastie T, Tibshirani R. Varying-coefficient models. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*. 1993; 55(B):757–779.
- Hedeker, D.; Gibbons, R. *Longitudinal data analysis*. John Wiley & Sons Inc.; Hoboken, NJ: 2006.
- Heron KE, Smyth JM. Ecological momentary interventions: Incorporating mobile technology into psychosocial and health behaviour treatments. *British Journal of Health Psychology*. 2010; 15(1): 1–39. doi: 10.1348/135910709X466063. [PubMed: 19646331]
- Honoré, PA.; Scott, W. *Priority areas for improvement of quality in public health*. Department of Health and Human Services; Washington, DC: 2010.
- Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*. 1998; 85:809–822. doi:10.1093/biomet/85.4.809.
- Hox, J.; Stoel, RD. Multilevel and SEM Approaches to Growth Curve Modeling.. In: Everitt, BS.; Howell, DC., editors. *Encyclopedia of Statistics in Behavioral Science*. Vol. 3. John Wiley & Sons, Inc.; Hoboken, NJ: 2005. p. 1296-1305.
- Huber PJ. The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. 1967; I:221–33.
- Hurvich CM, Tsai CL. Regression and time-series model selection in small samples. *Biometrika*. 1989; 76:297–307.
- Hussong AM, Cai L, Curran PJ, Flora DB, Chassin LA, Zucker RA. Disaggregating the distal, proximal, and time-varying effects of parent alcoholism on children's internalizing symptoms. *Journal of Abnormal Child Psychology*. 2008; 36(3):335. [PubMed: 17891557]
- Kaplan, DW. *Structural Equation Modeling: Foundations and Extensions*. 2nd Ed.. Sage Publications; Thousand Oaks, CA: 2009.

- Kenny PM, K. MT, Viney RC, Boyer MJ, Pollicino CA, McLean JM, Fulham MJ, McCaughan BC. Quality of life and survival in the 2 years after surgery for non small-cell lung cancer. *Journal of Clinical Oncology*. 2008; 26(2):233. [PubMed: 18086802]
- Lenderking WR, Hu M, Tennen H, Cappelleri JC, Petrie CD, Rush AJ. Daily process methodology for measuring earlier antidepressant response. *Contemporary Clinical Trials*. 2008; 29:867. [PubMed: 18606249]
- Li, R.; Root, T.; Shiffman, S. Models for intensive longitudinal data. Oxford University Press, Inc.; New York: 2006. A local linear estimation procedure for functional multilevel modeling.; p. 63
- Li WW, Lee TW, Yim AP. Quality of life after lung cancer resection. *Thoracic Surgery Clinics*. 2004; 14(3):353. [PubMed: 15382767]
- McArdle, JJ. Structural modeling experiments using multiple growth functions.. In: Ackerman, P.; Kanfer, R.; Cudeck, R., editors. *Learning and Individual Differences: Abilities, Motivation, and Methodology*. Erlbaum; Hillsdale, NJ: 1989. p. 71-117.
- McArdle JJ. Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*. 2009; 60:577–605.
- McArdle, JJ.; Bell, RQ. An introduction to latent growth curve models for developmental data analysis.. In: Little, TD.; Schnabel, KU.; Baumert, J., editors. *Modeling longitudinal and multiple-group data: practical issues, applied approaches, and scientific examples*. Erlbaum; Mahwah, NJ: 2000. p. 69-107.
- McArdle JJ, Epstein D. Latent growth curves within developmental structural equation models. *Child Development*. 1987; 58:110–133. [PubMed: 3816341]
- Meredith W, Tisak J. Latent curve analysis. *Psychometrika*. 1990; 55:107–122.
- Molenaar PCM, Sinclair KO, Rovine MJ, Ram N, Corneal SE. Analyzing developmental processes on an individual level using non-stationary time series modeling. *Developmental Psychology*. 2009; 45(1):260–271. [PubMed: 19210007]
- Munsch S, Meyer AH, Milenkovic N, Schlup B, Margraf J, Wilhelm FH. Ecological momentary assessment to evaluate cognitive-behavioral treatment for binge eating disorder. *International Journal of Eating Disorders*. 2009; 42(7):648. [PubMed: 19197978]
- Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine*. 2009; 39:1533. [PubMed: 19215626]
- Nagin, D. Group-based modeling of development. President and Fellows of Harvard College; United States of America: 2005.
- Portenoy RK, Thaler HT, Kornblith AB, Lepore JM, Friedlander-Klar H, Kiyasu E, Scher H. The memorial symptom assessment scale: An instrument for the evaluation of symptom prevalence, characteristics, and distress. *European Journal of Cancer*. 1994; 30(9):1326. [PubMed: 7999421]
- Preacher, KJ.; Wichman, AL.; MacCallum, RC.; Briggs, NE. Latent growth curve modeling. Vol. 157. Sage; 2008.
- Ruppert D. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*. 2002; 11:735–757. doi: 10.1198/106186002853.
- Schwartz G. Estimating the dimension of a model. *The Annals of Statistics*. 1978; 6(2):461.
- Schwartz JE, Stone AA. Strategies for analyzing ecological momentary assessment data. *Health Psychology*. 1998; 17(1):6. [PubMed: 9459065]
- Schwartz, JE.; Stone, AA. Stone, AA.; Shiffman, S.; Atienza, AA.; Nebeling, L., editors. *The analysis of real time momentary data: A practical guide.; The science of real-time data capture: Self-reports in health research*. 2007. p. 76
- Shiffman S. Dynamic influences on smoking relapse process. *Journal of Personality*. 2005; 73(6):1715. [PubMed: 16274451]
- Shiffman, S. Designing protocols for ecological momentary assessment.. In: Stone, AA.; Shiffman, S.; Atienza, AA.; Nebeling, L., editors. *The science of real-time data capture: Self-reports in health research*. Oxford University Press; New York: 2007. p. 27
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annual Review of Clinical Psychology*. 2008; 4:1.

- Shiffman S. How many cigarettes did you smoke? assessing cigarette consumption by global report, time-line follow-back, and ecological momentary assessment. *Health Psychology*. 2009; 28(5): 519–526. doi: 10.1037/a0015197. [PubMed: 19751076]
- Shiyko MP, Lanza ST, Tan X, Li R, Shiffman S. Using the time-varying effect model (TVEM) to examine dynamic associations between negative affect and self confidence on smoking urges: Differences between successful quitters and relapsers. *Prevention Science*. 2012; 13(3):288. [PubMed: 22246429]
- Shiyko MP, Li Y, Rindskopf D. Poisson growth mixture modeling of intensive longitudinal data: An application to smoking cessation behavior. *Structural Equation Modeling: A Multidisciplinary Journal*. 2012; 19(1):65–85. doi: 10.1080/10705511.2012.634722. [PubMed: 22408365]
- Shiyko MP, Naab P, Shiffman S, Li R. Modeling Complexity of EMA Data: Time-Varying Lagged Effects of Negative Affect on Smoking Urges for Subgroups of Nicotine Addiction. *Nicotine and Tobacco Research*. 2013
- Shiyko MP, Ram N. Conceptualizing and estimating process speed in studies employing ecological momentary assessment designs: A multilevel variance decomposition approach. *Multivariate Behavioral Research*. 2011; 46(6):875–899. doi: 10.1080/00273171.2011.625310. [PubMed: 22707796]
- Shrout, PE.; Lane, SP. Psychometrics.. In: Mehl, MR.; Conner, TS., editors. *Handbook of research methods for studying daily life*. Guilford Press; New York, NY: 2012. p. 302
- Singer, JD.; Willett, JB. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford University Press, Inc.; New York: 2003.
- Sohl SJ, Friedberg F. Memory for fatigue in chronic fatigue syndrome: Relationships to fatigue variability, catastrophizing, and negative affect. *Behavioral Medicine*. 2008; 34(1):29. [PubMed: 18400687]
- Stone AA, Schwartz JE, Broderick JE, Shiffman S. Variability of momentary pain predicts recall of weekly pain: A consequence of the peak (or salience) memory heuristic. *Pers Soc Psyc Bull*. 2005; 31:1340.
- Stone A, Shiffman S. Capturing momentary, self-report data: A proposal for reporting guidelines. *Annals of Behavioral Medicine*. 2002; 24(3):236–243. doi: 10.1207/s15324796abm2403_09. [PubMed: 12173681]
- Sugiura N. Further analysis of the data by Akaike's information criterion and the finite corrections. *Communications in Statistics - Theory and Methods*. 1978; 7:13–26.
- Tan X, Shiyko MP, Li R, Li Y, Dierker L. A time-varying effect model for intensive longitudinal data. *Psychological Methods*. 2012; 17(1):61. [PubMed: 22103434]
- U.S. Food and Drug Administration. [October 30, 2012] Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims. 2009. from <http://www.fda.gov/downloads/drugs/duidancecomplianceregulatoryinformation/guidances/ucm193282.pdf>
- Volkers AC, Tulen JHM, Van Den Broek Walter W, Bruijn JA, Passchier J, Pepplinkhuizen L. 24-hour motor activity after treatment with imipramine or fluvoxamine in major depressive disorder. *European Neuropsychopharmacology*. 2002; 12:273. [PubMed: 12126865]
- Walls TA, Schafer JL. *Models for intensive longitudinal data* Oxford University Press Inc. 2006
- Wand MP. Smoothing and mixed models. *Computational Statistics*. 2003; 18:223–249.
- Whalen CK, Henker B, Ishikawa SS, Emmerson NA, Swindle R, Johnston JA. Atomoxetine versus stimulants in the community treatment of children with ADHD: An electronic diary study. *Journal of Attention Disorders*. 2010; 13(4):391. [PubMed: 19474461]
- White H. Maximum likelihood estimation of misspecified models. *Econometrica*. 1982; 50:1–25.
- Wing RR, Papandonatos G, Fava JL, Gorin AA, Phelan S, McCaffery J, Tate DF. Maintaining large weight losses: The role of behavioral and psychological factors. *Journal of Consulting and Clinical Psychology*. 2008; 76(6):1015. [PubMed: 19045969]
- Yang, J.; Tan, X.; Li, R.; Wagner, A. TVEM (time-varying effect model) SAS macro suite users' guide (Version 2.1.0). The Methodology Center, Penn State; University Park: 2012. Retrieved from <http://methodology.psu.edu>

Figure 1.
Demonstration of the TVEM estimation approach with the spline method. A hypothetical function is split into seven sections by placing six equidistant knots (vertical dashed lines).

Figure 2.

Trajectories of symptom frequencies from 10 randomly selected lung-cancer patients (thin curves) recovering from either conventional (THOR) or video-assisted (VATS) surgical treatment. The overall group means (thick curves) are summaries for all patients ($N_{\text{THOR}} = 23$; $N_{\text{VATS}} = 36$).

Figure 3.

Trajectories of symptom burden from 10 randomly selected lung cancer patients (thin curves) recovering from conventional (THOR) or video-assisted (VATS) surgical treatments. The overall group means (thick curves) are summaries for all patients ($N_{\text{THOR}} = 23$; $N_{\text{VATS}} = 36$).

Panel A: estimated intercept function (# knots = 2) – a trajectory of symptom frequency over two weeks for conventional (THOR) group. Panel B: estimated slope function (# knots = 2; on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on symptom frequency over two weeks. Panel C: estimated group means for symptom frequency for THOR and VATS patients.

Figure 4.

Results of Generalized TVEM for symptom frequency scores (B-spline estimation).

Panel A: estimated intercept function (# knots = 2) – a trajectory of symptom frequency over two weeks for conventional (THOR) group. Panel B: estimated slope function (# knots = 2; on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on symptom frequency over two weeks. Panel C: estimated group means for symptom frequency for THOR and VATS patients.

Panel A: estimated intercept function – a trajectory of symptom frequency over two weeks for conventional (THOR) group. Panel B: estimated slope function (on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on symptom frequency over two weeks. Panel C: estimated group means for symptom frequency for THOR and VATS patients.

Figure 5.

Results of Generalized TVEM for symptom frequency scores (P-spline estimation).

Panel A: estimated intercept function – a trajectory of symptom frequency over two weeks for conventional (THOR) group. Panel B: estimated slope function (on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on symptom frequency over two weeks. Panel C: estimated group means for symptom frequency for THOR and VATS patients.

Panel A: estimated intercept function – a trajectory of symptom burden over two weeks for conventional (THOR) group. Panel B: estimated slope function (on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on the level of symptom burden over two weeks. Panel C: estimated group means for symptom burden for THOR and VATS patients.

Figure 6.

Results of Generalized TVEM for symptom burden scores (P-spline estimation).

Panel A: estimated intercept function – a trajectory of symptom burden over two weeks for conventional (THOR) group. Panel B: estimated slope function (on the risk ratio scale) – an effect of the video-assisted (VATS) treatment on the level of symptom burden over two weeks. Panel C: estimated group means for symptom burden for THOR and VATS patients.

Table 1

TVEM model selection for symptom frequency.

Number of knots for intercept & slope functions	AIC	BIC	AICC
Varying Intercept			
5,5	3983.7294	4071.5207	3984.4486
4,5	3981.9933	4064.9073	3982.6361
3,5	3980.4977	4058.5344	3981.0685
2,5	3979.6621	4052.8216	3980.1653
1,5	3980.8011	4049.0832	3981.2409
Varying Slope			
2,4	3980.5427	4048.8248	3980.9825
2,6	3981.2390	4059.2758	3981.8099
2,3	3979.4311	4042.8359	3979.8119
2,2	3978.8179	4037.3454	3979.1439
2,1	3981.2093	4034.8595	3981.4848

The final model is in **bold**.