spencer
health solutions

Estimating Quality of Life Trajectories from Intermittent At-Home Pulse Surveys

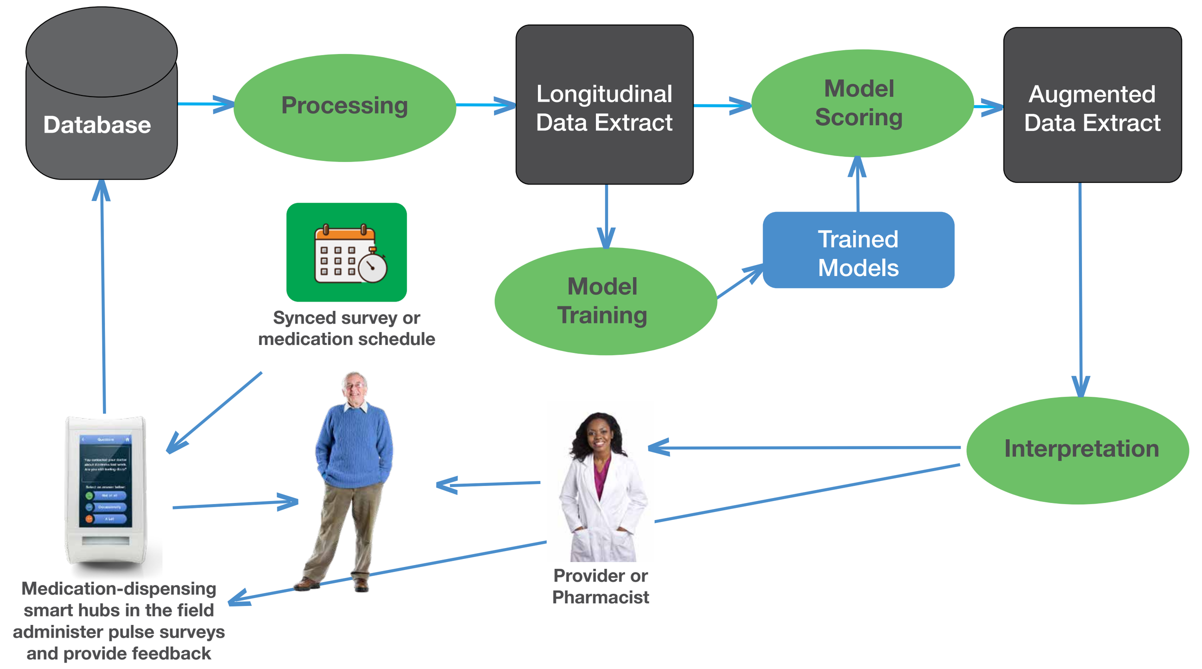
Ogorek BA, Kodavanti TD, Menius JA • Spencer Health Solutions, Morrisville, NC USA

Contact Ben Ogorek | bogorek@spencerhealthsolutions.com | 919-559-2649

Objectives & Background

Background

Understanding Quality of Life (QoL)[1] between clinical visits is challenging. Using a device that administers pulse surveys after dispensing medication pouches, we collect intermittent QoL data. The overall data generating process is described in the diagram below.



Complication: pulse survey data are sparse

Even when combining multiple QoL items, these data are sparse. In the example below, it takes two weeks to collect four Activity QoL responses.

date	are_you_more_or_less_active_today	have_you_exercised_today	planning_to_attend_a_social_activity_today
2/12/2022			
2/13/2022	0		
2/14/2022			
2/15/2022			
2/16/2022			-1
2/17/2022			
2/18/2022			
2/19/2022			
2/20/2022			1
2/21/2022		1	
2/22/2022			
2/23/2022			
2/24/2022			

A dynamic factor model handles sparsity & yields predictive QoL trajectories

The following dynamic factor model[2] has a latent state, estimable using either the Kalman Filter or Kalman Smoother[3], which we study as a time-varying clinical endpoint for QoL. The Gaussian distributions ensure tractable estimation, but constitute a model misspecification given our discrete data.

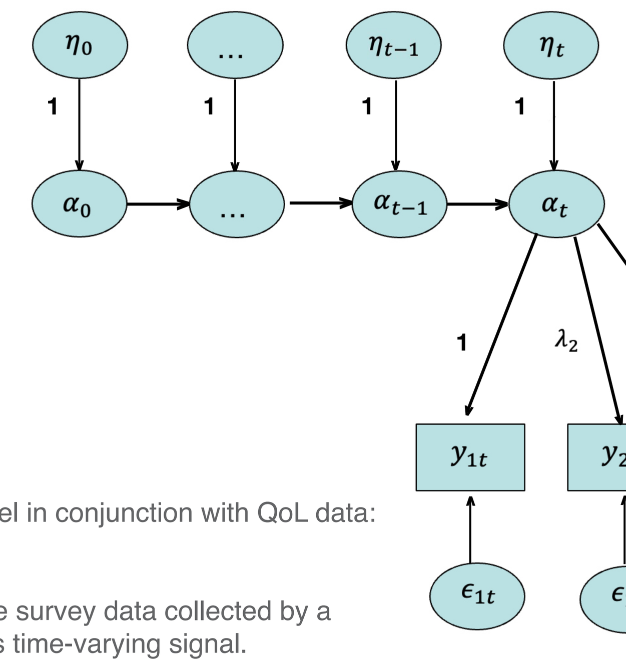
The Data (for one patient)
For questions i=1, 2, 3, and time t,

$$y_{it} \in \{-1,0,1,NA\}$$

The Model

$$\alpha_t = \alpha_{t-1} + \eta_t$$
$$y_{it} = \lambda_i \alpha_t + \epsilon_{it}$$
$$\alpha_0 \sim N(\mu_0, {\sigma_0}^2), \eta_t \sim N(0, \sigma_\eta^2), \epsilon_{it} \sim N(0, \sigma_{\epsilon_i}^2)$$

Graphical representation of the dynamic factor model for QoL



We can estimate the alpha sequence, called a "QoL trajectory," using either Kalman Filtering or Kalman Smoothing.

Using the estimates of the alphas from either the Kalman Filter or Kalman smoother, we can predict the values of the y_{it}.

Research objectives

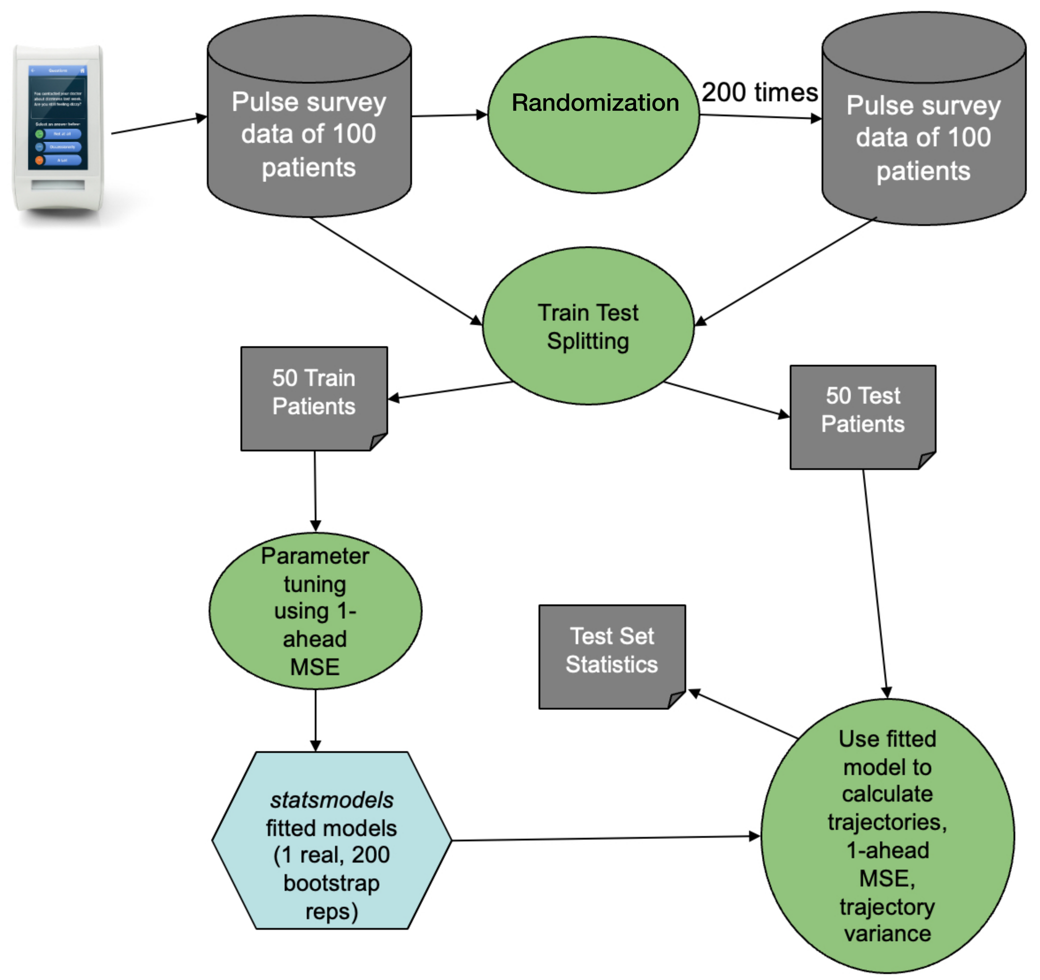
1. Study the feasibility of using the above dynamic factor model in conjunction with QoL data:
   a. For predicting future responses,
   b. For estimating a continuous QoL trajectory.
2. Use the dynamic factor model to demonstrate that the pulse survey data collected by a fleet of in-home medication-dispensing smart hubs contains time-varying signal.

Methods

Test the null hypothesis of random button pressing

The ability to reject the hypothesis of random button pressing is an initial bar for both model and data to jointly clear. We construct test statistics based on the following features of the model:

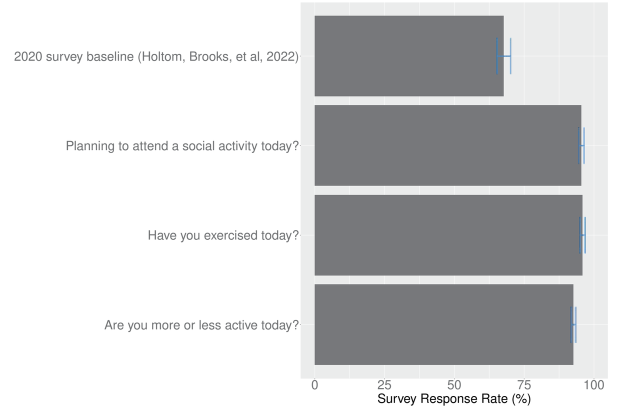
1. *The 1-ahead Mean Squared Error (MSE) for predicting future responses.* If the real data contain signal and the model is an effective representation, then we would expect the predictions from the model to have a lower 1-ahead MSE than for the random button pressing case.
2. *The Variance of the Smoothed Trajectories.* If the model can uncover real trajectories in the data and smooth noise from randomness, then the variance of the trajectories from the real data should exceed that of the random button pressing case.



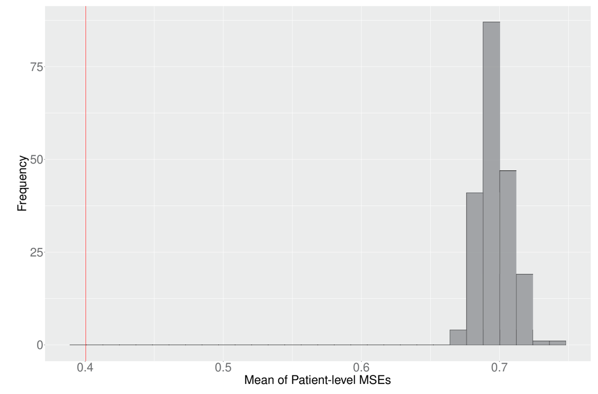
Procedural steps to test the hypothesis

1. Select 100 Canadian care management patients having answered at least 40 Activity QoL questions.
2. In 200 parametric bootstrap replicates, randomize responses to all real questions.
3. For all data sets, randomly assign 50 patients into the "train" group; the remaining 50 are assigned to the "test" group.
4. For the *train* patients in every data set, use the MLEmodel base class in the Python package statsmodels[4] to fit the dynamic factor model as follows:
   a. Set $\mu_0, {\sigma_0}^2$ to 0, 100^2 to reflect an uncertain initial state for each patient
   b. Initialize the following "tuning parameters": σ_η^2, and $\sigma_{\epsilon_i}^2$, λ_i, for i = 1, 2, 3.
   c. Use Kalman Filtering (via the MLEmodel's .filter method) to estimate the state at each pulse survey response. The state estimate enables prediction of each of the responses.
   d. Use the LBFGS optimization algorithm to minimize the 1-ahead MSE as a function of the tuning parameters.
5. For the *test* patients in every data set, use the fitted model objects to create QoL trajectories using both Kalman Smoothing and Kalman Filtering as follows:
   a. For 1-ahead MSE, use Kalman Filter trajectories with a seven question "burn in" period applied.
   b. For computing trajectory variance, use Kalman Smoother trajectories.
6. For both substeps of 5, test the null hypothesis of random button pressing by comparing the value obtained using the real data with the distribution of values obtained via the parametric bootstrap replicates.

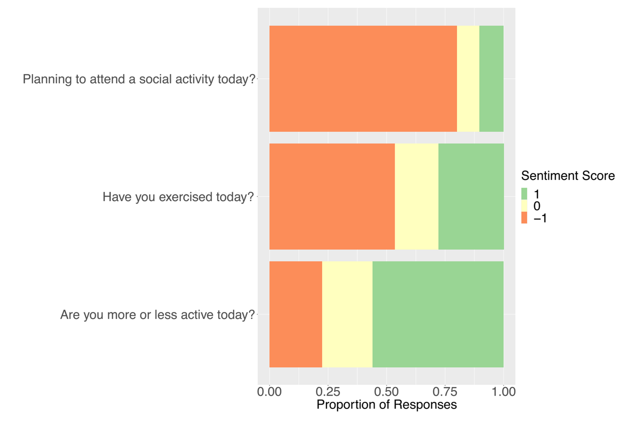
Results



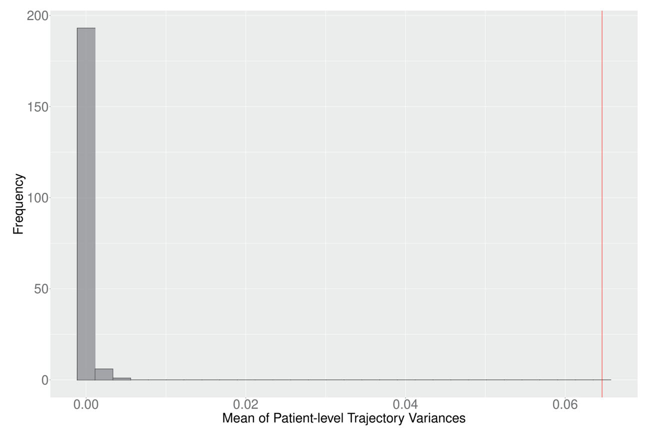
Pulse survey response rates were very high
For our cohort of 100 patients, 8860 Activity QoL questions were administered between 2017-10-29 and 2021-12-22 and 94.1% were answered. All three questions had a markedly higher response rate than the published 2020 baseline of 67.7%.



The random button pressing hypothesis was rejected using 1-ahead MSE from the dynamic factor model (via Kalman Filtering)
The mean of the 1-ahead MSE for predicting responses for the 200 simulation replicates was .697; for the real data it was .400, a relative decrease of 43%. As this value was smaller than all 200 bootstrap replicates, we rejected the null hypothesis of random button pressing at p < .005.

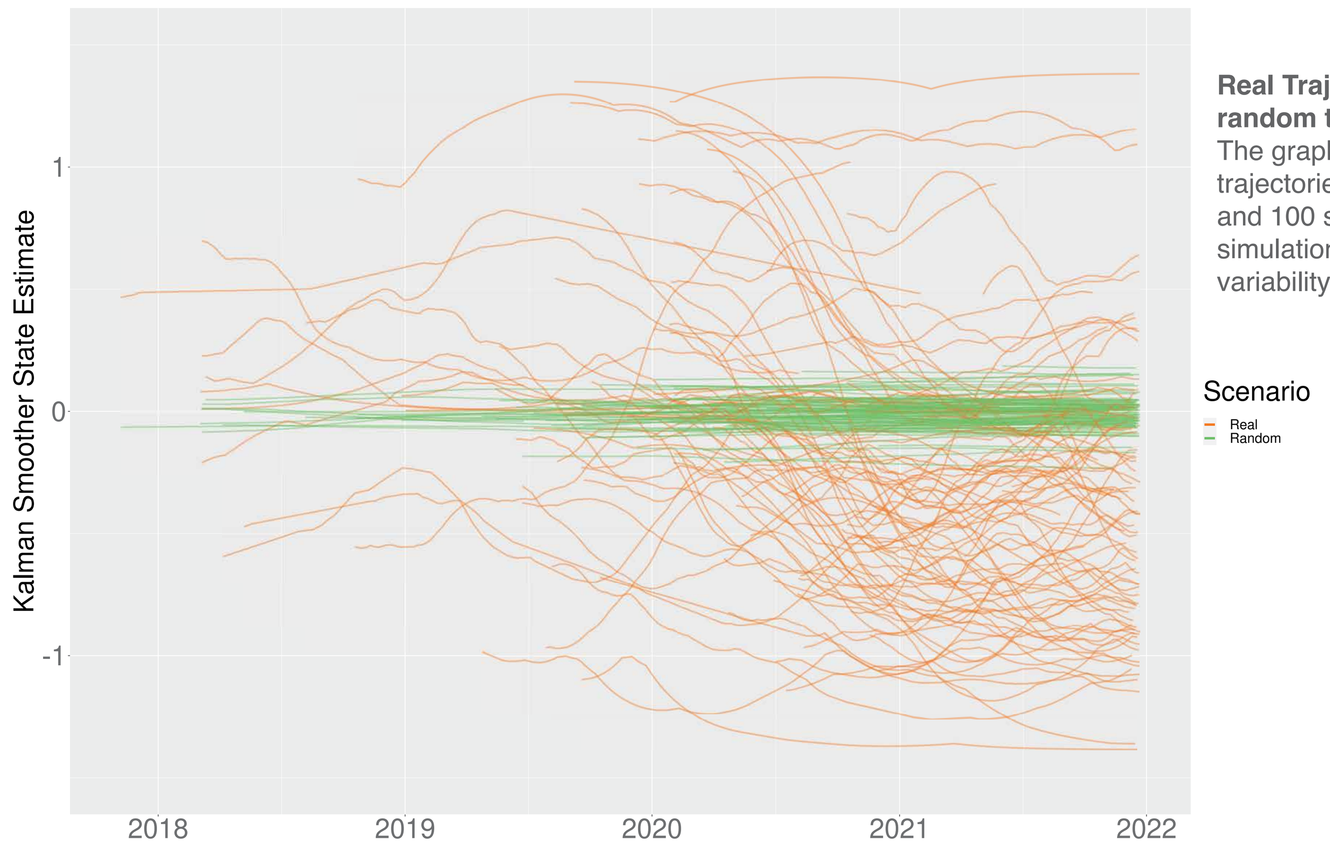


Response distributions differed by question
With the responses coded as -1, 0, and 1 to represent the positivity of the response sentiment, the questions had different patterns of responses, with *'Are you more or less active today?'* the most likely to earn a positive sentiment, and '*Planning to attend a social activity today?'* the least likely.



The random button pressing hypothesis was rejected using smoothed trajectory variances (via Kalman Smoothing)
The mean of the patient-level Kalman Smoother trajectory variances for the 200 simulated replicates was 1.67x10^-4; for the real data it was .065, or roughly 400x greater. As this value was larger than all 200 bootstrap replicates, we rejected the null hypothesis of random button pressing at p < .005.

Results (continued)



Real Trajectories looked different than random trajectories
The graph to the left depicts 100 smoothed trajectories using the actual responses (orange), and 100 smoothed trajectories using a simulation replicate (green). Note the reduced variability of the simulation trajectories.

Table: question-level statistics

Question	Times asked (Train and Test)	Response %	Factor Loading	Test Set responses after burn in	1-ahead MSE (SD)
Planning to attend a social activity today?	2036	95.5	1.00	608	.30 (.66)
Have you exercised today?	2051	96.0	.60	658	.63 (.60)
Are you more or less active today?	4773	92.7	.72	1925	.45 (.48)

Conclusions

Takeaways

- A survey delivered via medication-dispensing smart hub resulted in high response rates and differential response patterns.
- We were able to estimate continuous QoL trajectories using a dynamic factor model and a combination of Kalman Filtering and Smoothing.
- By rejecting the null hypothesis of random button pressing, we were able to show that these trajectories contained QoL signal, suggesting both a base-line level of data quality and methodological promise.
- The dynamic factor model had good properties despite a distributional misspecification; we used a 1-ahead predictive criteria to tune parameters.
- The model has predictive power for future responses. Questions with larger loadings had smaller mean squared errors.

Further work

- Compare the dynamic factor model to competing models and methods of aggregation.
- Validate trajectory trends against known outcomes.
- Incorporate this survey design and analysis into ongoing health outcome studies, with the QoL trajectory as a clinical endpoint.

References

1. Pequeno, N.P.F., Cabral, N.L.d., Marchioni, D.M. et al. Quality of life assessment instruments for adults: a systematic review of population-based studies. Health Qual Life Outcomes 18, 208 (2020). https://doi.org/10.1186/s12955-020-01347-7

2. Stock, James H., and Mark Watson. "Dynamic factor models." Oxford Handbooks Online (2011).

3. Markou, Stratis. "The Kalman Filter and Smoother." Random walks (n.d.). https://random-walks.org/content/misc/kalman/kalman.html.

4. Seabold, Skipper, and Josef Perktold. "Statsmodels: Econometric and statistical modeling with python."

5. Holtom, Brooks, et al. "Survey response rates: Trends and a validity assessment framework." human relations (2022): 00187267211070769.