# Longitudinal study of concussion-related diffusion MRI changes in college athletes

Nathan M. Muncy[1,*], Heather C. Bouchard[1], and Aron K. Barbey[1]

[1]Center for Brain, Behavior and Biology, University of Nebraska-Lincoln, Lincoln, Nebraska, USA

[*]Corresponding author. Email: nmuncy2@unl.edu

## Abstract

Sports-related traumatic brain injuries affect 1.6-3.8 million individuals in the US each year, and diffusion weighted imaging can measure the complex timeline of resulting axolemmal changes. Such longitudinal data is difficult to model statistically, however, given the high-dimensionality, semi-parametric and interdependent scalar values, and non-linear spatial (within-tract) and temporal (across visit) properties. Proposal: hierarchical generalized additive models (HGAMs) are well-suited to fit such data with the requisite flexibility and sensitivity to investigate (a) the spatial and temporal changes of white matter tracts, and (b) how such changes relate to diagnostic assessments. Methods: we utilized MRI and IMPACT data collected from 67 college athletes (9 female, age=19.43[1.68]) at three visits: start-of-season, post-concussion, and return-to-play. Diffusion tensors were modeled via constrained spherical deconvolution and probabilistic tractography from pyAFQ yielded 100 scalar values per white matter bundle. Results: By fitting the scalar profiles with longitudinal HGAMs we detected within-tract changes as a function of visit, revealing distinct patterns of post-injury disruption and recovery. Critically, it is unlikely that such changes would have been detected with standard techniques given their linear assumptions and limited dimensionality. Further, we examined whether these evolving diffusion metrics correlated with cognitive outcomes using HGAM tensor product interaction smooths and found moderate evidence linking white matter alterations to IMPACT composite scores. Merit: HGAMs offer a powerful framework to capture the complex progression of brain injury. Our findings suggest that HGAMs enhance our understanding of the spatiotemporal dynamics of brain injury and may enable more accurate tracking of injury and recovery.



KEYWORDS: DWI, MRI, GAM, TBI

# 1 Introduction

Introduction here.

# 2 Methods

## 2.1 Participants

Participants were recruited from men's football and women's soccer programs at the University of Nebraska-Lincoln, resulting in the enrollment of 69 (9 female, age = 19.36 $\pm$1.67, range = 17-24) National Collegiate Athletic Association (NCAA) athletes. Due to the limited number of females, and the sport-sex confound, we combined all participants into a single group. Institutional Review Board approval was obtained at the outset of the study, and prior to beginning experimental procedures participants completed informed consent and assent. Magnetic Resonant Imaging (MRI) and clinical assessment (ImPACT) data were acquired during three sessions: enrollment at the beginning of the season (baseline, Base), within 48 hours of diagnosed concussion (post-concussion, Post), and prior to return-to-play (RTP). As MRI and ImPACT (below) data were gathered separately, a number of participants did not contribute MRI and/or ImPACT data across one or more of the sessions. This resulted in the following final session counts: Base = 67 MRI (9 female), 61 ImPACT (5 female), Post = 65 MRI (8 female), 48 ImPACT (3 female), and RTP = 56 MRI (7 female), 32 ImPACT (2 female).

## 2.2 ImPACT

Description of ImPACT.

## 2.3 MRI Protocol

Magnetic Resonance Imaging data were collected on a 3-Tesla Siemens MAGNETOM Skyra scanner at the Center for Brain, Behavior and Biology (University of Nebraska-Lincoln) utilizing a 32-channel coil. For each of three sessions (Base, Post, and RTP), participants contributed T1 and diffusion weighted images (T1w, DWI). T1w Multi-Echo Magnetization Prepared - RApid GRadient Echo (MEMP-RAGE) structural scans were acquired with the following parameters: TR = 2530 ms, TE = 1.69, 3.55, 5.41, and 7.27 ms, flip angle = 7°, voxel size = 1 $mm^3$, FoV = 256 × 256, slices = 176 interleaved. DWI scans were acquired via TR = 3000 ms, TE = 95 ms, flip angle = 90°, voxel size = 1.719 × 1.719 × 2.4 $mm^3$, 134 slices, multi-band acceleration factor = 3, directions = 128, bandwidth = 1500 Hz/Px, shells = 1 (b-value = 1000 $s/mm^2$), reference volumes = 6 (b-values = 0 $s/mm^2$; $b_0$). A set of field maps for the DWI scans were collected using the same acquisition direction (anterior-posterior; AP) and reversed (posterior-anterior; PA).

## 2.4 MRI Data Processing

Preprocessing and modeling of the DWI data were conducted using FSL v6.0 (Jenkinson et al., 2012) and PyAFQ v1.3.6 (Kruper et al., 2021; Yeatman et al., 2012). First, $b_0$ volumes from AP and PA field map files were extracted and combined, as were their acquisition parameters. Next, `topup` calculated a distortion correction matrix from the AP-PA $b_0$ file. A brain mask was constructed via `bet`, and an index file was generated to describe the relationship between the DWI volumes and their acquisition parameters. Preprocessing of DWI was then conducted via `eddy_openmp`, thereby producing motion- and distortion-corrected diffusion images.

Whole-brain tractography was computed from the preprocessed DWI by PyAFQ. Constrained spherical deconvolution was used to derive the fiber orientation distribution function (fODF) of each voxel, where constrained-positivity regularization = 1, minimum amplitude $\tau = 0.1$, mean gray matter diffusivity = 0.0008, mean CSF diffusivity = 0.003, 600 fODF

iterations, and spherical harmonics order = 8. Resulting fODFs of each voxel were then utilized to probabilistically generate fiber maps, using one seed per voxel for each dimension, a maximum turning angle of 30°, step size = 0.5 mm, and a length range = 50-250 mm. The resulting fibers were parcellated into individual tracts via *a priori* inclusion (waypoint) and exclusion regions of interest (Wakana et al., 2007). These tracts were then compared to a fiber probability map (Hua et al., 2008) and any fibers which traverse low-probability spaces were removed from the tract. Further, any fibers with a length 3+ standard deviations from the tract average, or 4+ standard deviations from the average path centroid, were removed as well. Lastly, each tract was then resampled into 100 equidistant nodes (according to a Mahalanobis distance metric) from which averaged diffusion values and scalars were calculated. Specifically, for each tract node we extracted averaged axial diffusivity ($\lambda_{\parallel}$; AD), radial diffusivity ($(\lambda_{\perp 1} + \lambda_{\perp 2})/2$; RD), mean diffusivity ($(\lambda_{\parallel} + \lambda_{\perp 1} + \lambda_{\perp 2})/3$; MD), and fractional anisotropy (FA).

## 2.5 GAM specification

Generalized additive models (GAM) are an extension of general linear models capable of modeling high-dimensional data which contain non-linear relationships. Where regression models fit data with a linear (or higher-order polynomial) function, GAMs construct a smooth curve to fit data from a set of basis functions (i.e. splines). Such a smooth can capture complex X-Y relationships that would be underfit by models with linear assumptions. Further, high dimensional relationships can be modeled via 3-dimensional smooths (i.e. membrane), termed a 'tensor product interaction smooth', or with hypersurfaces for higher dimensions (Baayen & Linke, 2020). Such capabilities have made GAMs useful in fields such as ecology ([CITE]) and linguistics ([CITE]), which often model complex data in high dimensions or across multiple factors, and researchers using MRI techniques are beginning to adopt the method ([CITE]). We recently demonstrated their applicability to modeling DWI scalar data (Muncy et al., 2022), and here we extend GAMs to model high-dimensional,

longitudinal, multimodal data.

Hierarchical GAMs (HGAMs; Pedersen et al., 2019) allow for model fits at both global and group levels. That is, it is possible to model both the X-Y relationship that is shared across all levels of a factor (global smooth) and differences that factor levels (group smooths) may have from the global smooth. Further, it is not required that each level of smooth (global, group) contain the same 'wiggliness' in the X-Y relationships. Separate smooth curves and wiggliness terms at different factor levels of HGAMs is highly relevant in modeling concussion-related changes within white matter tracts, as the global smooth of the tractometric profile (i.e. scalar values across all nodes) can effectively be held constant when modeling potential changes across session, and independent wiggliness terms may capture scalar changes unique to one time point. Further, tensor product interaction terms can be utilized to build multimodal models, investigating the relationship of the tractometric profile with independent metrics such as the ImPACT composite scores. Accordingly, such a model would be capable not only of detecting changes within a tract that result from concussion, but also how such changes relate to clinical assessments. Finally, and critically, HGAMs facilitate conducting longitudinal, whole-brain analyses on tractometric profiles as data from all tracts and across all time points can be included in the same model. Such a specification allows for within-subject pooling of variance across both tract and time. Where modeling individual tracts results in a creeping Type-I error and the corresponding corrections, injury (and subsequent recovery) may affect multiple tracts within a subject and such shared variance would be lost when investigating tracts individually. By including all tracts and time points, HGAMs have the capability to not only reduce Type-I but also Type-II errors. All GAMs were specified using the `mgcv` package version 1.9-1 (Wood, 2017) in R version 4.3.3 (R Core Team, 2023).

### 2.5.1 Longitudinal difference model

To investigate within-tract injury- and recovery-related FA changes we specified an HGAM to test for Post and RTP tract FA differences from Base. First, we calculated the Post-Base and RTP-Base changes in FA ($\Delta$FA). While including original FA values would be ideal, propagating ordered factors (Base < Post < RTP) across an interaction with another factor (tract) loses the original ordered structure; ordered factors would be necessary to investigate differences from baseline instead of merely the interaction with session. Next, we calculated the session comparison $\times$ tract interaction term as `mgcv::bam` does not currently support modeling smooths by factor interactions. $\Delta$FA values were modeled as a function of tract node using thin-plate regression splines (R Code 1) and a basis dimensionality of 15 was determined sufficient to fit the tract curves (`gam.check(fit_LDI)`). Subjects were treated as a random effect, thereby allowing each subject to have their own intercept across all levels of the factors, the $\Delta$FA distribution was well-fit by a Gaussian distribution with an identity link function, fast Residual Error of Maximum Likelihood (fREML) was used as the smoothing parameter estimation method, and 12 threads were used in the computation (run time $\approx$ 45 minutes). Input data consisted of the 24 tracts with good segmentation across all subjects. Notably, we did not include a global smooth for this model, as the $\Delta$FA profile would differ for each tract, and we specified that each tract would have its own wiggliness term; essentially this is a longitudinal model of FA differences which references model 'I' in Pedersen et al. (2019).

```
fit_LDI <- mgcv::bam(
  delta_fa ~ s(subj_id, by=tract_scan, bs="re") +
    s(node_id, by=tract_scan, bs="tp", k=15) +
    tract_name+sess_comp+tract_scan,
  data=df,
  family=gaussian(),
  method="fREML",
  nthreads=12
)
```

R Code 1: ΔFA values are modeled as a function of tract node with thin-plate regression smooths for each tract, accounting for the within-subject factors of tract and session and using separate wiggliness terms for each tract. `delta_fa` = RTP-Base and Post-Base FA differences, `subj_id` = subject identifier factor, `node_id` = node identifier integer, `tract_name` = tract identifier factor, `sess_comp` = session comparison factor (RTP-Base, Post-Base), and `tract_scan` = interaction of `tract_name` and `sess_comp`.

### 2.5.2 Longitudinal tract model

The model specified in R Code 1 effectively models the entire longitudinal dataset of ΔFA values, allowing for pooling for variance within a subject across tract and session, not requiring a multiple comparison correction for modeling all tracts. But as the ΔFA calculation required data at time points A and B, the analysis was restricted by missing data. As essentially a post-hoc analysis to further interrogate tract differences across session, and also what change in scalar (e.g. increased RD) drove the difference in FA, individual tracts were modeled with a longitudinal HGAM with terms for global and group smooths (R Code 2). Tract FA values were fit by a beta distribution with a logit link function, AD and RD values were fit with a Gaussian distribution and identity link function, and a gamma distribution with a logit link function fit the MD values. Subjects were again treated as a random effect, with separate intercepts for each scan (Base, Post, RTP), group smooths were allowed their own wiggliness parameter, and the colinearity of global and group smooths was controlled

by the 'm' parameter. Such a model is similar to model 'GI' in Pedersen et al. (2019). Additionally, converting the session factor to an ordered factor was used in a separate model to test for differences in Post and RTP scalar values from Base (Supplemental R Code 4). Such a model is particularly useful as the test statistic, which describes the flatness of the smooth, provides information about changes from Base values rather than deflections from zero.

```
fit_LGI <- mgcv::bam(
  <scalar> ~ s(subj_id, scan_name, bs="re") +
    s(node_id, bs="tp", k=15, m=2) +
    s(node_id, by=scan_name, bs="tp", k=15, m=1),
  data=df,
  family=<family>,
  method="fREML",
  nthreads=4
)
```

R Code 2: Tract scalars are modeled as a function of tract node with thin-plate regression splines using both global and group (`scan_name`) smooths as well as individual group wiggliness. `<scalar>` = relevant DWI metric (AD, RD, MD, or FA), `scan_name` = session identifier factor (Base, Post, RTP), `<family>` = relevant family and link function for scalar distribution.

### 2.5.3 Longitudinal tract interaction model

As noted above, GAMs are capable of modeling higher-dimensional, non-linear interactions through tensor product interaction smooths and hypersurfaces, a property which make them particularly relevant for multimodal research. We used such a model to test whether concussion- and recovery-related changes in tract scalars related to changes in ImPACT composite and total symptom scores (R code 3), thereby potentially linking damage within a specific region of a tract to changes in assessment metrics. Tract scalars were modeled as a

function of both tract node and ImPACT measure, and the node-ImPACT interaction term was specified such that each session (Base, Post, RTP) would have a different scalar-node-ImpACT interaction surface. We note the decrease in basis dimensionality for the ImPACT measures thin-plate regression splines from the default value, and that fitting the tensor product interaction smooth also benefited from a slightly higher basis dimensions term for the tract node term. Finally, a model using ordered factors was also specified to derive a test statistic against Base rather than zero (Supplemental R Code 5).

```
fit_LGI_intx <- mgcv::bam(
  <scalar> ~ s(subj_id, scan_name, bs="re") +
    s(node_id, bs="tp", k=15, m=2) +
    s(imp_meas, by=scan_name, bs="tp", k=5) +
    ti(
      node_id, imp_meas, by=scan_name,
      bs=c("tp","tp"), k=c(20,5), m=1
    ),
  data=df,
  family=<family>,
  method="fREML",
  nthreads=4
)
```

R Code 3: Tract scalars are modeled as a function of separate 2D node and ImPACT smooths as well as a 3D tensor product interaction surface. `imp_meas` = ImPACT composite or total symptom measure.

### 2.5.4 ImPACT model

The relationship between session (Base, Post, RTP) and ImPACT composite metrics (verbal memory, visual memory, visual motor, impulse control, and reaction time) and total symptom scores were modeled with GAMs to test for changes across assessment session. As with

tract scalar profiles, GAMs were employed as (a) non-linear trends are expected in such metrics, and (b) they can model the semi-parametric distributions encountered in several of the metrics. Each ImPACT metric was fit as a function of assessment number, using integer values rather than categorical Base, Post, and RTP (Supplemental R Code 6); such a specification allowed for modeling evolving changes in assessment metrics rather than comparing main effects across factor levels. Verbal and visual memory composites were converted to proportion scores and modeled with a beta distribution and logit link function, visual motor and reaction time were best fit with Gaussian distributions and identity link functions (despite the skewness), and a negative binomial distribution with log link function fit the impulse control and total symptoms well.

When specifying models, whether with ImPACT or DWI data, model fits were reviewed and assessed via `mgcv:gam.check()`, and the selection of competing models was aided by `itsadug::compareML()`. Pipeline and statistical code, information about their respective environments, and curated data are available at the project repository: https://github.com/nmuncy/adr_dwi.

# 3 Results

## 3.1 ImPACT

ImPACT assessment smooths (Section 2.5.4) were exctracted and plotted for visualization purposes (Figure 1). All models except for impulse control detected a significant interaction between ImPACT metric and assessment number. Visual memory, reaction time, and total symptoms had patterns consistent with concussion-related deficits at Post and subsequent recovery at RTP (visual memory: $F_{(1.94,1.99)} = 8.59$, $p < .001$; reaction time: $F_{(1.91,1.99)} = 6.18$, $p < .01$; total symptoms: $F_{(1.98,1.99)} = 28.74$, $p < .0001$). We also note that total symptoms at RTP were much lower than at Base (Figure 1, bottom right). Conversely, while verbal memory and visual motor tests indicate significant non-flatness (verbal memory: $F_{(1.82,1.96)} =$

4.34, $p = .028$; visual motor: $F_{(1.86,1.97)} = 8.19$, $p < .001$), their values did not differ between Base and Post while RTP scores were significantly better. This pattern possibly reflects a lack of sensitivity at Base and/or practice effects. Finally, impulse control was unchanged (i.e. flat) as a function of assessment ($F_{(1.0,1)} = .003$, $p = .95$).

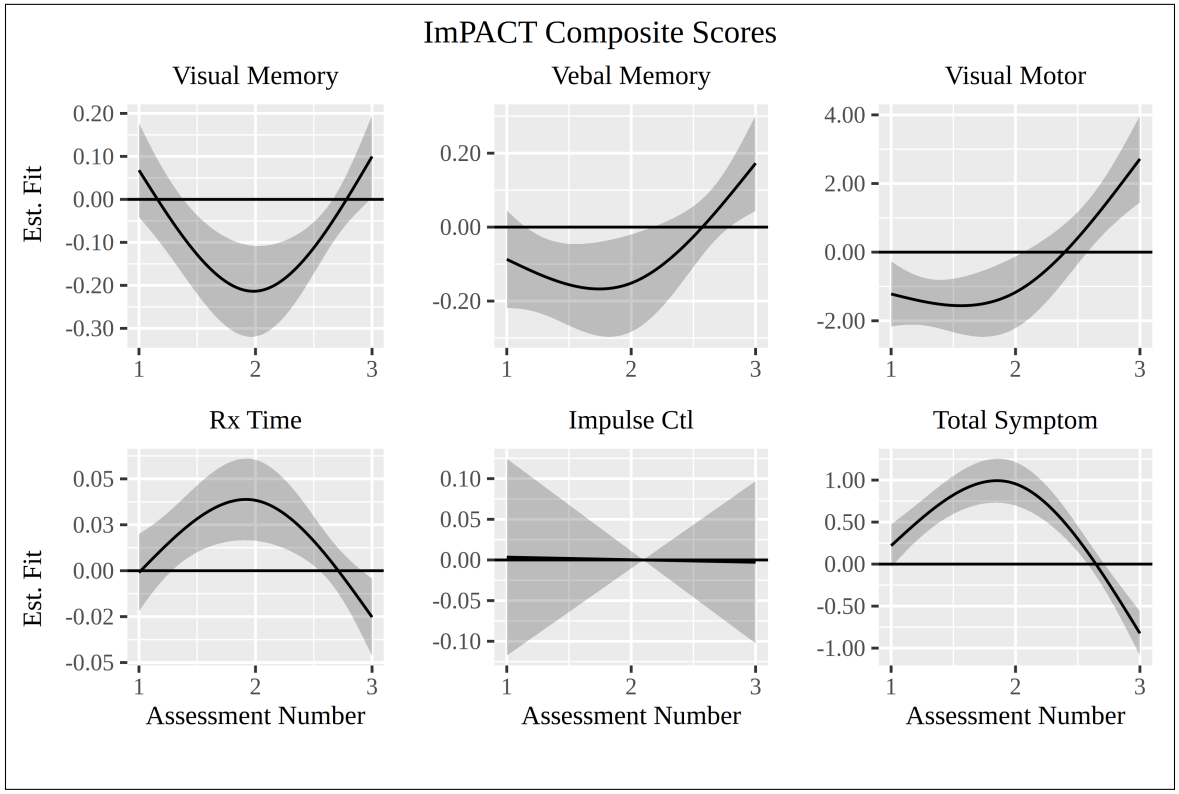


Figure 1: GAM smooths for ImPACT composite and total symptom scores. Assessment numbers where the confidence interval does not include 0 indicate significant changes. Visual memory, reaction time, and total symptoms showed worsening and then recovery (U-shapes) while verbal memory and visual motor scores were better at assessment 3. Impulse control did not change across assessments. Assessment number 1=Base, 2=Post, 3=RTP. Rx Time = reaction time, Impulse Ctl = impulse control.

## 3.2 DWI Tracts

Tract results.

## 3.3 DWI Tracts Interactions - ImPACT

Description of DWI - ImPACT interaction.

### 3.4 DWI Tracts Interactions - Time

Description of DWI-time interaction.

## 4 Discussion

Discussion.

## Acknowledgments

People. Grant.

# References


Baayen, R. H., & Linke, M. (2020). An introduction to the generalized additive model. *A practical handbook of corpus linguistics*, 563–591.

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., Calabresi, P. A., Pekar, J. J., van Zijl, P. C., & Mori, S. (2008). Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification. *Neuroimage*, *39*(1), 336–347.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). Fsl. *NeuroImage*, *62*(2), 782–790.

Kruper, J., Yeatman, J. D., Richie-Halford, A., Bloom, D., Grotheer, M., Caffarra, S., Kiar, G., Karipidis, I. I., Roy, E., Chandio, B. Q., Garyfallidis, E., & Rokem, A. (2021). Evaluating the Reliability of Human Brain White Matter Tractometry. *Aperture neuro*, *1*(1), 10.52294/e6198273-b8e3-4b63-babb–6e6b0da10669. https://doi.org/10.52294/e6198273-b8e3-4b63-babb-6e6b0da10669

Muncy, N. M., Kimbler, A., Hedges-Muncy, A. M., McMakin, D. L., & Mattfeld, A. T. (2022). General additive models address statistical issues in diffusion MRI: An example with clinically anxious adolescents. *NeuroImage: Clinical*, *33*, 102937. https://doi.org/10.1016/j.nicl.2022.102937

Pedersen, E. J., Miller, D. L., Simpson, G. L., & Ross, N. (2019). Hierarchical generalized additive models in ecology: An introduction with mgcv. *PeerJ*, *7*, e6876. https://doi.org/10.7717/peerj.6876

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. Manual. R Foundation for Statistical Computing. Vienna, Austria.

Wakana, S., Caprihan, A., Panzenboeck, M. M., Fallon, J. H., Perry, M., Gollub, R. L., Hua, K., Zhang, J., Jiang, H., & Dubey, P. (2007). Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage*, *36*(3), 630–644.

Wood, S. N. (2017). *Generalized additive models: An introduction with R*. CRC press.

Yeatman, J. D., Dougherty, R. F., Myall, N. J., Wandell, B. A., & Feldman, H. M. (2012). Tract Profiles of White Matter Properties: Automating Fiber-Tract Quantification. *PLOS ONE*, *7*(11), e49790. https://doi.org/10.1371/journal.pone.0049790

# 5 Supplemental Materials

Supplemental Materials.

```
df$scanOF <- factor(df$scan_name, ordered=T)
fit_LGIO <- mgcv::bam(
  <scalar> ~ s(subj_id, scan_name, bs="re") +
    s(node_id, bs="tp", k=15, m=2) +
    s(node_id, by=scanOF, bs="tp", k=15, m=1),
  data=df,
  family=<family>,
  method="fREML",
  nthreads=4
)
```

R Code 4: Tract scalars are modeled as a function of tract node with thin-plate regression splines using both global and group smooths, individual wiggliness terms for groups, and ordered factors to compare Post and RTP group smooths to Base.

```
df$scanOF <- factor(df$scan_name, ordered=T)
fit_LGIO_intx <- mgcv::bam(
  <scalar> ~ s(subj_id, scan_name, bs="re") +
    s(node_id, bs="tp", k=15, m=2) +
    s(imp_meas, by=scan_name, bs="tp", k=5) +
    ti(node_id, imp_meas, bs=c"tp","tp"), k=c(20,5), m=1) +
    ti(
      node_id, imp_meas, by=scanOF,
      bs=c("tp","tp"), k=c(20,5), m=1
    ),
  data=df,
  family=<family>,
  method="fREML",
  nthreads=4
)
```

R Code 5: Tract scalars are modeled as a function of separate 2D node and ImPACT smooths as well as a 3D tensor product interaction surface, with ordered factors used to compare Post and RTP surfaces to Base.

```
fit_G <- mgcv::bam(
  imp_meas ~ s(subj_id, bs="re") +
    s(num_assess, bs="tp", k=3),
  data=df,
  family=<family>,
  method="fREML"
)
```

R Code 6: ImPACT metrics modeled as a function of number of assessments using a single global smooth. `imp_meas` = ImPACT composite or total symptom score, `num_assess` = assessment number (1=Base, 2=Post, 3=RTP).

## 5.1 Tables

Supplemental Tables.

## 5.2 Figures

Supplemental Figures.