SARA WESTON
Assistant Professor
University of Oregon
sweston2@uoregon.edu

JENNIFER PFEIFER
Professor
University of Oregon
jpfeifer@uoregon.edu

MICHELLE BYRNE
Senior Lecturer
Monash University
michelle.byrne@monash.edu

Secondary data analysis & pre-registration of longitudinal analyses

ABCD Workshop on Brain Development and Mental Health

Sponsored by NIMH

June 22nd - July 22nd, 2021

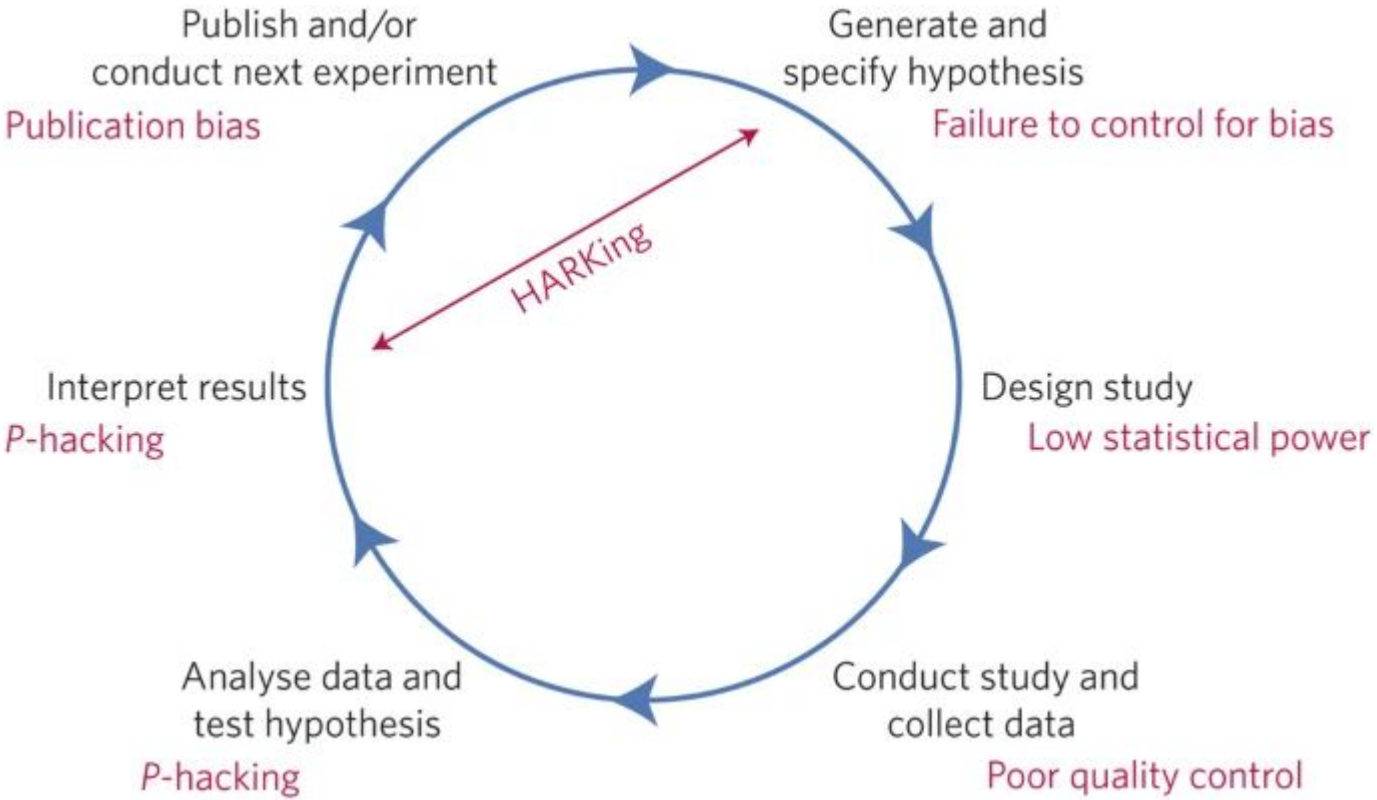
What is a preregistration?

A document outlining the **planned** analyses.

- i.e., we preregister before we start playing with data.

Why?

- We develop hypotheses before we analyze the data.
- Preregistrations serve as a document of what our expectations and plans were.
- Identify which analyses are based on theory and which are based on the data.



From Munafò et al. (2017).
Preregistration can help guard against issues of *p*-hacking, HARKing, and publication bias.

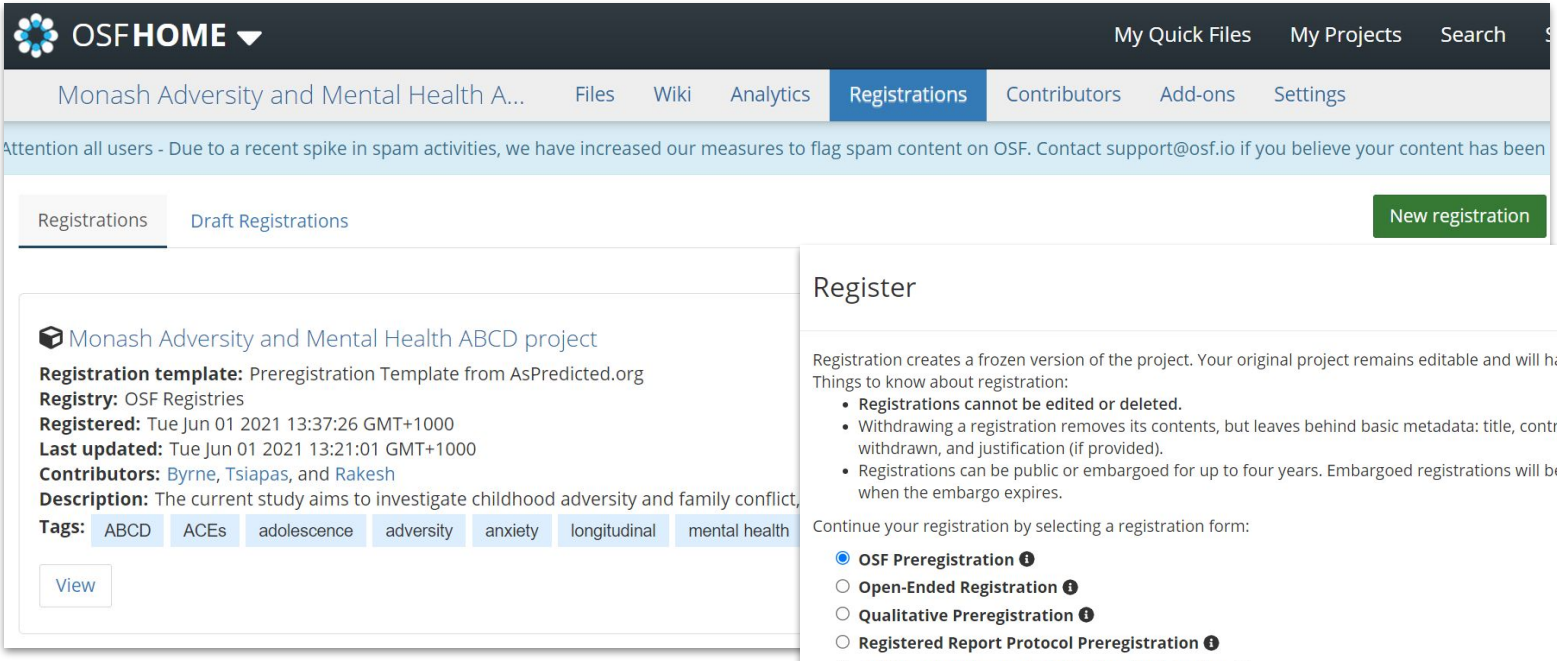
But don’t do it for science…

… do it for you!

- Focus on the research question.
- Increase credibility in your results.
- Stake your claim on an idea.
- Not a waste of time.
    - A reordering of steps.

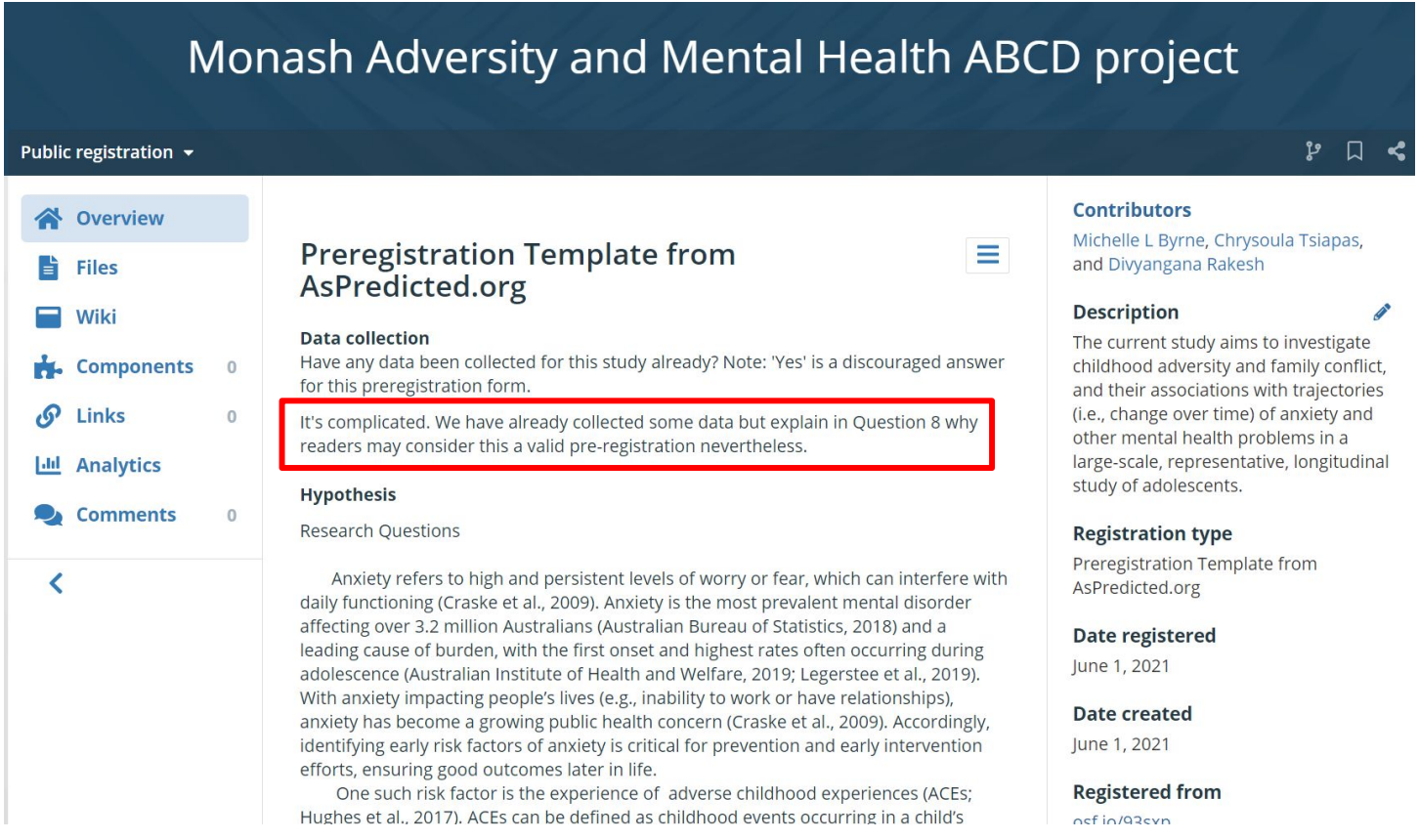
Photo by airfocus on Unsplash

Components of a preregistration - Example



https://osf.io/cjnh4

Components of a preregistration



https://osf.io/cjnh4

Components of a preregistration

Dependent variable

This study will use data collected from the Adolescent Brain Cognitive Development (ABCD) study (https://abcdstudy.org/).

The dependent variable will be scores of anxiety symptoms as measured by the Child Behavior Checklist (CBCL) and the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS) for DSM-5 (KSADS-5).

Conditions

How many and which conditions will participants be assigned to?

N/A

Analyses

This study will use a "model building sample" of participants with only two waves of data: baseline and 1-year follow-up (n ~ 4664) separated into three subsamples (training, holdout 1 and holdout 2) of approximately 1500 participants each.

1. An Exploratory Factor Analysis (EFA) model of ACEs indicators will explore the latent structure of ACEs using the training sample.

2. The first Confirmatory Factor Analysis (CFA) will use the same latent structure from the EFA to explore model fit and refine in the holdout 1 sample

Outliers and Exclusions

The sample size may end up being reduced from the "model building" sample of 4664 & "experimental" sample of 6571. Our analyses will include anyone who has at least one observed ACEs measure, i.e., we will not exclude listwise participants who are missing only some ACEs measures. We will report the number of participants in each sample who were missing all ACE measures and they will not be included in this study. However, because we do not yet know the frequency and patterns of missingness on ACEs data, we may choose to conduct post hoc sensitivity analyses for participants who are missing >50% of ACEs measures. If our models do not converge potentially due to high rates of missing ACEs data, we may also choose to have a higher threshold for excluding listwise and conduct analyses again.

Sample Size

The ABCD dataset has over 11000 participants available. This study intends to use all those that have data available (see above).

Other

The dataset is publicly available. Although access to ABCD data has been granted, no analyses have been conducted.

Name

Monash Adversity and Mental Health ABCD project

Finally

Observational/archival study

Other

No response

https://osf.io/vefsz

Components of a preregistration - fMRI study

TAG Sharing Task

Summary

Provide a narrative summary of what is contained in this registration, or how it differs from prior registrations.

Project working title: Behavioral and neural phenotypes of self-disclosure during adolescence.

The current study. This study investigated the behavioral and neural phenotypes associated with disclosure of self-referential information to peers during early adolescence (10.0 – 13.1 years). Participants underwent a functional magnetic resonance imaging (fMRI) scan while completing a modified version of the "pay-per-view" or monetary choice task. Specifically, they made decisions about whether they would disclose superficial and intimate information about themselves (i.e. "share" vs. "keep private") to a chosen peer (ideally their best friend). Each disclosure decision was associated with a monetary value, and the reward associated with self-disclosure was quantified as the amount of money that adolescents would forgo to share information. Aside from this behavioral index, we examined the neural underpinnings of intimate and superficial self-disclosure. We also examined how these behavioral and neural phenotypes related to the quality of peer relationships, self-perceptions and mental health.

Contributors
Nandita Vijayakumar

Registration type
Open-Ended Registration

Date registered
June 18, 2018

Date created
June 18, 2018

Registered from
osf.io/qx9d8

Category
Project

Citation
osf.io/vefsz

perceived social competence and self-worth (SPPA: close friendship, social competence, global self-worth).

2. Neuroimaging aims
Examine neural activation associated with self-disclosure and how this differs based on statement depth (intimate vs. superficial). It is hypothesized that :
a) sharing information (relative to keeping it private) will be associated with greater activation in regions subserving reward, social and self-referential processes.
b) intimate disclosure will recruit the same networks to a greater extent than superficial disclosure, in addition to regions involved in affective and regulatory processes.
c) activation will be modulated by monetary reward, such that greater monetary incentives for disclosure will increase activation of the same regions during intimate vs. superficial disclosure.

Examine whether individual differences in neural activation relate to socioemotional functioning. Both modulated and un-modulated brain activation will be related to friendship quality and support, and perceived social competence and self-worth. It is hypothesized that:
d) greater neural activation of the social cognitive and reward network during intimate vs. superficial disclosure will be related to better friendship quality and support.
e) greater neural activation of the social cognitive and reward network during intimate vs. superficial disclosure will be related to greater perceived social competence and global self-worth.

Method

Participants
A community sample of 182 adolescent girls, aged 10.0 to 13.0 years (mean = 11.56 years, SD = 0.81 years), were recruited from Lane County, Oregon, into a longitudinal project called Transitions in Adolescent Girls (TAG). The majority of participants were recruited from primary and middle schools in the region, and a small subset were

https://osf.io/vefsz

Components of a preregistration - fMRI study

fMRI data acquisition

Data was acquired on a 3T Siemens Skyra MRI scanner at the Lewis Center for Neuroimaging at the University of Oregon. High-resolution T1-weighted structural images were collected with the MP-RAGE sequence (TE=3.41 ms, TR=2500 ms, flip angle=7°, 1.0 mm slice thickness, matrix size=256 x 256, FOV=256 mm, 176 slices, bandwidth=190 Hz/pixel). Two functional runs of T2*-weighted BOLD-EPI images were acquired with a gradient echo sequence (TE=27 ms, TR=2000 ms, flip angle = 90°, 2.0 mm slice thickness, matrix size=100 x 100, FOV=200mm, 72 slices, bandwidth=1786 Hz/pixel). There were 60 to 87 images per run, as run length varied with participants' response times during Cyberball. To correct for local magnetic field inhomogeneities, a field map was also collected (TE=4.37 ms, TR=639.0 ms, flip angle=60°, 2.0 mm slice thickness, matrix size=100 x 100, FOV=200 mm, 72 slices, bandwidth=1515 Hz/pixel).

Behavioral measures of interest

Friendship intimacy

Intimate Friendship Scale (IFS) examines the quality of friendships across eight dimensions: "frankness and spontaneity", "sensitivity and knowing", "attachment", "exclusiveness", "giving and sharing", "imposition", "common activities", and "trust and loyalty". The questionnaire consists of 32 items that are rated on a seven-point Likert scale ranging from "always disagree" to "always agree". Internal reliability (alphas) of the subscales range from 0.72 to 0.77 (Sharabany et al., 1974).

fMRI

Raw DICOM image files were converted to the NifTI format with MRIConvert (Smith, 2011) and organized according to the Brain Imaging Data Structure (BIDS) standards (Gorgolewski et al., 2016), which facilitates the use of portable analysis tools called BIDS Apps (Gorgolewski et al., 2017). fMRI data were then preprocessed using the fmriprep BIDS App (v1.0.0; https://github.com/poldracklab/fmriprep; Esteban et al., 2017), a tool based on Nipype (Gorgolewski et al., 2011). Each T1-weighted (T1w) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrectionv2.1.0 [4] and skull-stripped using antsBrainExtraction.sh v2.1.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6 [5], and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle [20]. Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c [6] was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 [7], using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast [16] (FSL v5.0.9). Functional data was motion corrected using mcflirt (FSL v5.0.9 [8]). Distortion correction was performed using fieldmaps processed with fugue [11] (FSL v5.0.9). In the case of 1 participant who was missing fieldmaps, "Fieldmap-less" distortion correction was performed by co-registering the functional image to the same-subject T1w image with intensity inverted [12,13] constrained with an average fieldmap template [14], implemented with antsRegistration (ANTs). This was followed by co-registration to the corresponding T1w using boundary-based registration [15] with 9 degrees of freedom, using bbregister (FreeSurfer v6.0.1). Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using antsApplyTransforms (ANTs v2.1.0) using Lanczos interpolation. Frame-wise displacement [18] was calculated for each functional run using the implementation of Nipype. Many internal operations of FMRIPREP use Nilearn [21], principally within the BOLD-processing workflow. For more details of the pipeline see https://fmriprep.readthedocs.io/en/latest/workflows.html. Following FMRIPREP, pre-processed output will be smoothed in SPM12 using a kernel of 6-mm FWHM Gaussian kernel.

https://osf.io/vefsz

Components of a preregistration - fMRI study

Whole brain analyses
Aim 2a & 2b: A repeated-measures flexible factorial ANOVA with a 2x2 design will examine the interaction between the statement depth (Intimate, Superficial) and disclosure (Share, Private) factors. We will specifically examine whether disclosure differed across the two statement depth conditions, and whether intimate disclosure differed from superficial disclosure. Given the inclusion of the parametric modulator of shareValue in the first (subject)-level models, interpretation of the unmodulated response is the mean activation across all trials (i.e. controlling for ShareValue).

Aim 2c: A repeated-measures flexible factorial ANOVA with a 2x2 design will examine the interaction between the shareValue parametric modulators for statement depth (Intimate, Superficial) and disclosure (Share, Private) factors. As above, we will specifically examine the main effect of disclosure, and whether intimate disclosure differed from superficial disclosure.

Whole brain analyses will be corrected for multiple comparison using cluster-wise correction, by employing AFNI 3dClustSim. Smoothness estimates entered into 3dClustSim will be spatial autocorrelation function (acf) parameters averaged from each individual's group level model residuals, as calculated by 3dFWHMx using the -acf flag.

ROI analyses
Activation in independently defined ROIs of the reward and social cognitive system will be examined. A mask of the vmPFC will be created with Neurosynth (http://www.neurosynth.org), using reverse inference with the search term "mentalizing". 6 mm spheres will be created around the peak z-values of this map. Anatomical masks of the left and right VS will be extracted from the Harvard-Oxford subcortical atlas, with the threshold of 25%. Contrast estimates for the main conditions of interest will be extracted for each of these three ROIs.

Aim 2a & 2b: A repeated-measures ANOVA with a 2x2 design will examine the interaction between the statement depth (Intimate, Superficial) and disclosure (Share, Private) factors.

Aim 2c: A repeated-measures ANOVA with a 2x2 design will examine the interaction between the shareValue parametric modulators for statement depth (Intimate, Superficial) and disclosure (Share, Private) factors.

For each aim, Bonferroni correction for multiple comparisons will be undertaken that accounts for the mean correlation between the three ROIs (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm)

Do I have to make my preregistration public?

Public preregistration

- Good citizen of science
  - Change the cultural norm
  - Share troublespots so others can learn
  - No more file drawer!
- Reap rewards
  - Badges at journals
  - Findings may be perceived as more trustworthy

Private preregistration

- Maintains all personal benefits
  - Self-trust in results
  - Better planning
  - Better interpretation
- Protect your idea
  - Although, idea stealing is rare
    Embargo periods
  - If this is what you want, create a form and save it on your computer.

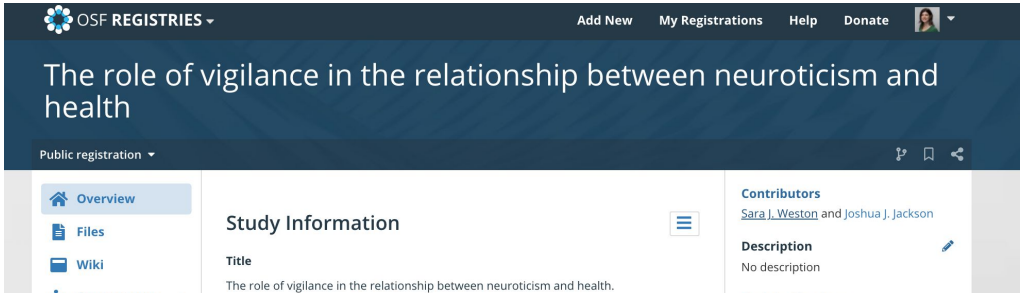


Photo by Eliott Reyna on Unsplash

Unique challenges with pre-existing data

Possibility of prior exposure to the data

- Remember, a goal of preregistration is to separate the analytic decisions made based on theory and those made based on data.
- Exposure can be subtle.

Have you worked with these data before?

- Report prior analyses -- even unpublished

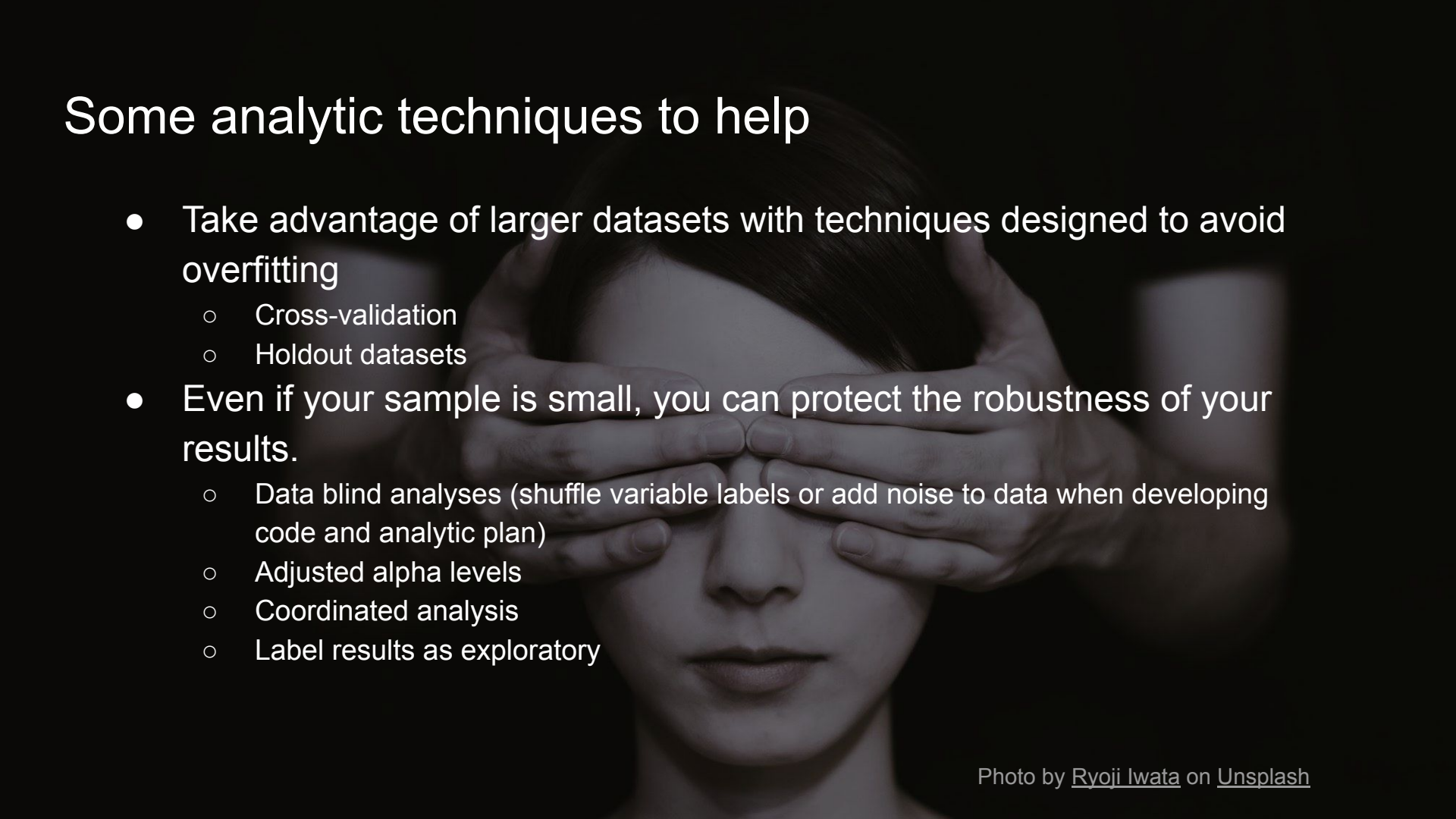
Have you **read about** these data before?

- Report what you've read.

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science, 2*(3), 214-227.

Some analytic techniques to help

- Take advantage of larger datasets with techniques designed to avoid overfitting
  - Cross-validation
  - Holdout datasets
- Even if your sample is small, you can protect the robustness of your results.
  - Data blind analyses (shuffle variable labels or add noise to data when developing code and analytic plan)
  - Adjusted alpha levels
  - Coordinated analysis
  - Label results as exploratory



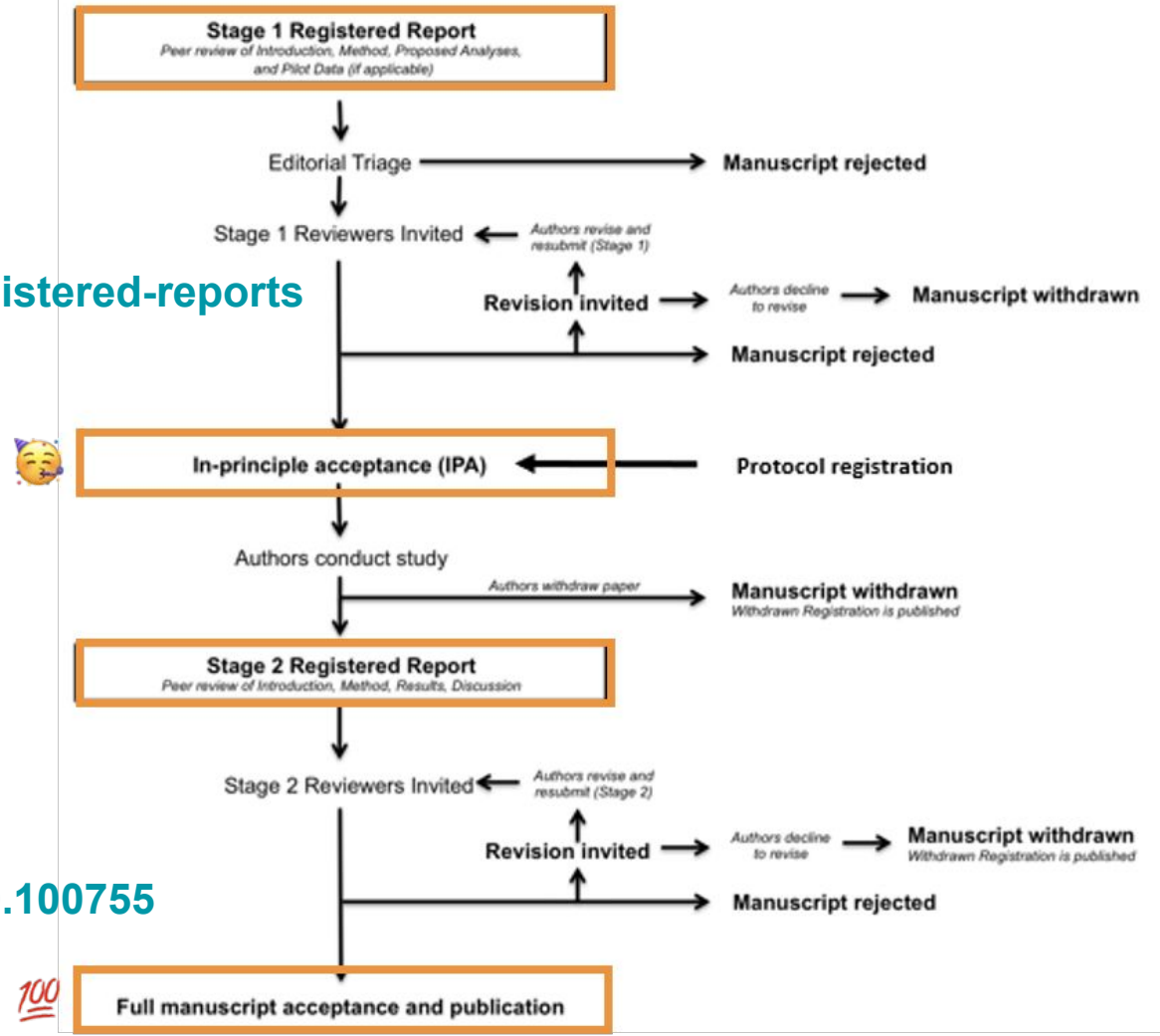
Registered Reports

https://www.cos.io/initiatives/registered-reports

Project **undergoes peer review** prior to data collection (‘primary’) or data analysis (‘secondary’)

Introduction to this in context of developmental cognitive neuroscience (Pfeifer & Weston, 2020)

https://doi.org/10.1016/j.dcn.2020.100755



Where can I submit Registered Reports?

Developmental Cognitive Neuroscience

Cortex

Journal of Cognitive Neuroscience

Neuroimage: Reports

Developmental Science

Infant & Child Development

Peer Community in Registered Reports https://rr.peercommunityin.org/about/about

Registered Report Example

Barrett et al., in prep

What access did you ALREADY have to the data?

Methods

Sample Characteristics

This project will use three timepoints of data gathered through the Transitions in Adolescent Girls* (TAG) study. The third wave of data collection was just over halfway completed before being interrupted by the COVID-19 pandemic, and we will only use data collected before the pandemic for these analyses. Analyses to date have not yet examined change in brain structure. Only one analysis has inspected measurements of structural brain data within this sample. The first wave of structural scans was processed and analyzed for a separate study on the association between pubertal hormone concentration and cortical thickness and subcortical volume (Vijayakumar, Pfiefer, Flournoy, Hernandez, & Dapretto, 2019). Out of the adversity questionnaires, only socioeconomic status measurements have been used in prior analyses (specifically an examination of self-perceived scholastic competence across adolescence). The proposed analyses herein have not yet been conducted and are not influenced by existing analyses that have examined other research questions, which have inspected neither structural brain data nor measurements of adversity.

189 adolescents aged 10.00 to 13.00 years and their caregivers were recruited from the Lane County community in Oregon, USA primarily through recruitment letters issued through schools (all districts in the greater Eugene/Springfield area) and to a smaller extent through

How will you test your hypotheses? What will you take as evidence of support?

Hypotheses

1. It is hypothesized that each pattern of similarity in adversity, as defined by both the cumulative risk model and the DMAP (threat and deprivation), will be significant predictors of similarity patterns in structural brain development within the prefrontal, parietal, and medial temporal lobe networks. Specifically, similarity in cumulative risk will significantly predict similarity in structural brain development (model 1); also, similarity in threat and similarity in deprivation will together significantly predict similarity in structural brain development (model 2). We further hypothesize that both DMAP predictors (similarity in threat and similarity in deprivation) in the model 2 equation will have significant beta coefficients.
2. A comparison of fit statistics between the two models is hypothesized to indicate that using two similarity patterns to represent threat and deprivation separately will predict more variance in trajectories of structural brain development than similarity along one cumulative adversity pattern.

To test the hypotheses, we will build dissimilarity matrices of both adversity-related experiences and structural trajectories among individuals. Dissimilarity values will be calculated using Euclidean distances between participants along the measures of interest. Adversity values will be assembled into an RDM for each dimension (threat, deprivation, cumulation); the height and width of which will correspond to the number of participants and will be symmetric around the diagonal. Each cell will contain a distance value between two participants. Similarly, a structural brain change RDM will represent participants' distances along their annualized percent change values in each specified brain region.

These RDMs will then be vectorized above the diagonal, and the adversity RDMs will be included in regression models used to predict the structural change RDM. To test the first hypothesis, the structural change RDM will be regressed along the threat and deprivation RDMs (structural change RDM ~ threat RDM + deprivation RDM). All effect sizes and fit statistics will be reported. To test the second, the structural change RDM will be regressed along the cumulative experiences RDM (structural change RDM ~ cumulation RDM). The conceptual models will be compared using their Akaike information criterion (AIC) to determine which provides the best parsimonious fit of the data. We will use the rules of thumb provided by Burnham & Anderson (2004) to determine if there is substantial support for the DMAP in comparison to a cumulative risk-based model. If the DMAP outperforms the cumulative risk model, an item-level cumulative risk RDM will be regressed along the structural change RDM and compared to the DMAP to determine whether performance differences are solely related to a difference in variance within the matrices. Tests will be considered statistically significant at an alpha level of 0.05.

ROIs

The following regions of interest (ROIs) will be examined in this study: Amygdala, hippocampus, rostral anterior cingulate cortex, medial orbital frontal cortex, rostral middle frontal cortex, and superior parietal cortex. We will use anatomical masks based on the Harvard-Oxford subcortical atlas that have been created within our lab for the amygdala and hippocampus (Vijayakumar et al., 2019). Cortical regions will be defined by selecting the appropriate parcellations from the Desikan-Killiany atlas given its use in the relevant literature.Cortical regions will be defined by selecting the appropriate parcellations from the Desikan-Killiany atlas given its use in the relevant literature.

To address the concerns of the CTQ physical neglect subscale's psychometric properties, we plan to conduct a sensitivity analysis with physical neglect items and subscale score in each RDM to determine whether inclusion alters the significance of the predictors. Before creating the RDMs, group comparisons between participants who are excluded due to poor scan quality and participants who are included will be conducted for all adversity measurements (CTQ scores, traumatic events score, parental income, bullying, and free lunch) to see whether data loss is greater among individuals with more adverse exposures. Measurements of internal consistency for all CTQ subscales will be reported for included participants.

Statistical Power Analysis

After assessing scan quality, 112 participants are eligible for these analyses from the first wave. While quality control is still underway for the SOS substudy and subsequent waves, it is reasonable to expect that approximately 100 participants will have usable scans from at least two time points. Previous studies have reported small to medium effects of adverse experiences on neural development observed in smaller samples (Gold et al., 2016; McLaughlin

Being specific in defining ROIs/COIs etc. is essential! (and difficult)

Contingency plans - "what happens when…"

Exploratory Analyses?

Can be discussed at Stage 1 where they are necessary to justify study variables or procedures that are included in the design exclusively for exploratory analysis.

Planned Exploratory Analyses

The following planned exploratory analyses will not be included in Step 1 submission of this registered report, but are allowed in Step 2. In order to assess the adequacy of each conceptual model, we will flip the prediction models to use similarities in neural regions to predict similarities in past experiences. To test the dimensional model, an RDM of threat experiences will be regressed on an RDM of threat-implicated ROIs (those within the fronto-amygdala network and hippocampus), an RDM of deprivation-implicated ROIs (those within the frontoparietal network), and an RDM with all ROIs. The predictors will be compared using a Mantel test that accounts for interdependence among the matrices. The same procedure will be repeated to predict similarities in deprivational experiences. To predict similarities along a cumulative experience RDM, we will include an RDM with all established ROIs and an RDM representing similarities in structural change within the whole brain. Each predictor will again be compared using a Mantel test. Assessing the relative utility of each predictor allows us to perform a similarity “mapping” of neural clusters to experiences. This method will test which similarities among different neural groupings are strongest predictors of similarities in participants’ history of adverse experiences, assessing the claim within the DMAP that threat and deprivation result in altered development in uniquely specified ROIs. Using the cumulative risk model, we will also examine whether similarities in ROIs traditionally associated with cognitive and emotional processing differences for individuals with a history of childhood adversity or similarities in whole brain development are more predictive of similarities in all adverse experiences.

Timeline

If Stage 1 review is successful, since the data is already collected, the next step will be to inspect the remaining scans for motion and surface quality, and then complete the analyses. The authors expect to complete that process and resubmit within a subsequent three-month period.

Question Time / Breakout Rooms

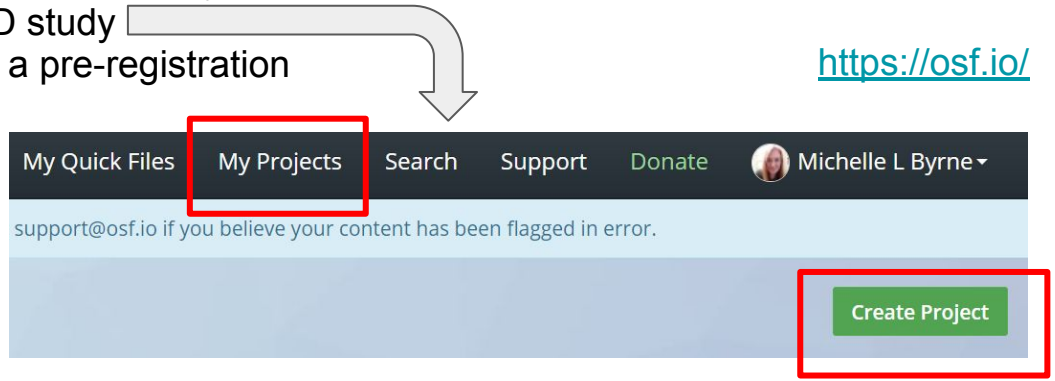
There is time now to ask questions and think about your own preregistration! There is no need to complete these and/or submit them yet, but you can discuss with your breakout room colleagues how you plan to organize your preregistration, and even start drafting your own. We encourage you to share screens!

We will move between breakout rooms to answer questions.

Things you can do now (and we will help you!):

- Create an OSF account/profile if you don't already have one
- Create an OSF project for your ABCD study
- Choose a template and start drafting a pre-registration
- Start drafting a registered report

https://osf.io/



Further Resources

Workshop on Meaningful Effects:
https://apps1.seiservices.com/meaningfuleffects/09022020_MeaningfulEffects_Summary.pdf

Evaluating Effect Size in Psychological Research: Sense and Nonsense, Funder & Ozer 2019:
https://journals.sagepub.com/doi/full/10.1177/2515245919847202

Example of a behavioural ABCD preregistration:
https://osf.io/cjnh4

fMRI Preregistration Template by Dr Jessica Flannery, from the 2018 Brainhack:
https://osf.io/6juft/

Improving practices and inferences in developmental cognitive neuroscience, by Dr John Flournoy et al:
https://www.sciencedirect.com/science/article/pii/S1878929320300554

Example of a neuroimaging pregregistration from the longitudinal TAG study, by Dr Nandi Vijayakumar:
https://osf.io/vefsz
Paper published from that prereg:
https://pubmed.ncbi.nlm.nih.gov/32039615/

Instructions and tutorials for OSF prereg:
https://help.osf.io/hc/en-us/articles/360042097853-Create-a-View-only-Link-for-a-Registration

Recommendations for Increasing the Transparency of Analysis of Preexisting Data Sets by Dr Sara Weston et al:
https://pubmed.ncbi.nlm.nih.gov/32190814/

Example of corrections/deviations from prereg:
https://osf.io/xmb9w/