This is the submitted version of a paper accepted at *Brain, Behavior, and Immunity*. Please cite as “A primer on common analytic concerns in psychoneuroimmunology: Alternatives and paths forward. Brain, Behavior, and Immunity, 102, 338-340. https://doi.org/10.1016/j.bbi.2022.03.007”

A Primer on Common Analytic Concerns in Psychoneuroimmunology: Alternatives and Paths Forward

Daniel P. Moriarity[1,2]

[1]Department of Psychology, Temple University

[2]Department of Psychiatry, McLean Hospital/Harvard Medical School

Correspondence concerning this article should be addressed to Daniel P. Moriarity, Department of Psychology, Temple University, Weiss Hall, 1701 N. 13th St., Philadelphia, PA 19122, United States of America. E-mail: Daniel.moriarity@temple.edu.

Funding: Daniel P. Moriarity was supported by National Research Service Award F31MH122116 and an APF Visionary Grant.

Psychoneuroimmunology has been an increasingly popular area of research in recent years, as evidenced by the number of submissions to *Brain, Behavior, and Immunity* (the premier journal in the field) increasing from 873 in 2016 to 2,491 in 2020. As this field has grown so has interest in methodological work aiming to maximize the replicability and clinical impact of this research. For instance, recent work has begun to characterize the physiometrics (measurement properties of biological variables; Moriarity and Alloy, 2021; Segerstrom and Smith, 2012) of inflammatory variables. For example, some studies have investigated the temporal stability of inflammatory proteins (Out et al., 2012; Shields et al., 2019) and the temporal specificity of effects between immune and psychological variables (Moriarity et al., 2019). Others have tested how many biological samples are necessary to achieve specific analytic goals (e.g., reliability, Segerstrom and Boggero, 2020; Shields et al., 2019). Further, some recent work in this journal has investigated whether removing observations with CRP>10 mg/L (often cited as an indicator of acute illness) is an appropriate default strategy (Mac Giollabhui et al., 2020; Moriarity et al., 2021b). Other researchers have leveraged designs and analytic techniques that isolate theoretically-relevant variance (e.g., within-person variance; Kuhlman et al., 2018; Moriarity et al., 2020) or investigate inflammation in the context of established psychosocial models of psychiatric risk (e.g., rumination; Moriarity et al., 2018; Szabo et al., 2022). However, the replicability and clinical impact of this work is contingent upon statistical rigor and proper execution of analyses.

This current manuscript seeks to add to these bodies of work by briefly, and accessibly, discussing analytic concerns commonly observed in psychoneuroimmunology research and introduce alternatives. It is worth note that the issues described below are not specific to psychoneuroimmunology; however, I hope that this viewpoint might serve as a resource for

methodological reform in the subfield. To maximize the ease of integrating the below recommendations into new work, supplemental materials includes template code for all suggestions and links to additional resources (e.g., further reading, Shiny apps).

Linear Regressions Assume Normality of Residuals *Not* Values

Many psychoneuroimmunology studies investigate hypothesized linear relationships between immunological and psychosocial/behavioral variables. For example, whether higher levels of CRP are associated with higher levels of depression or whether individuals who perseverate more during a stressor have greater inflammatory reactions. Consequently, linear regressions (and other types of linear models) are commonly used in psychoneuroimmunology. A common misconception in linear regression is that the assumption of normality refers to the distribution of the data (Ernst and Albers, 2017), which is often seen in psychoneuroimmunology research (see Horn et al. (2018) for an example meta-analysis that reports how often studies of the relation between CRP and depression report normality of CRP data and how it influenced the statistics used). While this is true for some types of analyses (e.g., bivariate correlations (Kowalski, 1972)), the assumption of normality for linear regressions (and many other types of models) refers to the distribution of the residuals (i.e., error terms; Tabachnick and Fidell, 2013). In other words, the majority of cases should have small residuals/error, with fewer and fewer cases as error increases. This is because error is used to calculate estimates; thus, violation of this assumption could result in biased estimates and/or p-values (Ernst and Albers, 2017). R code to investigate normality of residuals in linear regression has been provided in Supplemental Materials using data from the 2017-2020 NHANES cohort (Centers for Disease Control and Prevention, 2009). Should residuals be skewed, the most commonly suggested remedial measure

to try transforming the dependent variable; however, sometimes it is possible to use a statistical approach robust to non-normality (described below).

Some readers might be concerned that, because log-transformation is commonly treated as a default method in many psychoneuroimmunology studies, increasing the frequency of raw biomarkers as variables would introduce difficulties for future meta-analyses. This should not be an issue because accounting for potential differences in effect sizes when proteins are raw vs. log-transformed could be accounted for by testing transformation-status as a moderator. Further, the quality of meta-analyses are limited by the methodology of their component studies; thus, advancing methodological rigor is one of the most direct ways to improve the quality of meta-analyses.

Sometimes it is Best to Change Models, Not Data

It is common for data to be manipulated via transformation and winsorization/removal of outliers, given the oft skewed nature of immunological data (and many psychological variables of interest) and frequency of extreme values. Although these decisions might help standard models run smoothly, certain types of data manipulation result in data that no longer fully represents the original sample. For example, consider a sample of adolescents without autoimmune conditions recruited to investigate the relationship between IL-6 and depression. If a subset of observations with elevated IL-6 are removed or winsorized to satisfy statistical assumptions, this decision effectively adds a third inclusion criterion (i.e., adolescent, no autoimmune conditions, and IL-6 below a specific level). Additionally, it is important to consider whether inflammatory outliers might represent individuals particularly relevant to the theories under study (e.g., hypotheses that atypically elevated inflammation is a risk factor for psychopathology; Mac Giollabhui et al., 2020; Moriarity et al., 2021b). Further, individuals with

atypically high inflammation may naturally be more common in populations of interest (e.g., those with chronic illness or acute stress). It is also common for datasets to have observations below the lower limit of detection, which are often deleted or imputed to a specific value (discussed in greater detail in Horn et al., 2018).

In addition to the conceptual implications of data manipulation, there can be quantitative and interpretive consequences as well. For example, deleting observations might result in biased estimates and can directly reduce power by decreasing sample size. Further, the ability to make substantive interpretations can also decrease after manipulating data. Consider a linear regression testing the relationship between CRP (measured in mg/L) and depression symptoms resulting in a coefficient of .5. This would be interpreted as every 1 mg/L increase in CRP predicts a .5 increase in the depression score, which is more easily interpretable at a practical level than an estimate in log(mg/L) and/or log depression score. Additionally, analyses of log-variables results in multiplicative effects instead of additive effects (in other words, proportional vs. summative effects; Liu and Maxwell, 2020). This is not inherently incorrect (and might be suitable to much psychoneuroimmunology research) but should be considered when deciding whether to transform data.

To clarify, this is not a condemnation of transformation, winsorization, or outlier removal. Rather it is an invitation to consider whether there are benefits to keeping data true to their observed form (e.g., interpretability, retaining extreme values that represent populations of interest) and explore alternatives (Courvoisier and Renaud, 2010). For example, if the assumption of normality is violated there might be nonparametric alternatives (e.g., using the Kruskal-Wallis test instead of a one way ANOVA). To avoid outliers on one's dependent variable from overly influencing the results of a regression, robust regression could be used.

Similarly, there are a number of robust estimators available for latent variable modeling (e.g., MLR). Code for these alternatives is included in Supplemental Materials.

Biological Composites Need to be Physiometrically Evaluated Before Use

Biological aggregates are becoming increasingly popular in psychoneuroimmunology to counteract issues with multiple comparison and selective reporting associated with analyzing multiple analytes. Additionally, if hypotheses are not specific to the effects of specific proteins (e.g., inflammation generally), a latent or composite variable might better match the level at which the underlying theory is described. When composites are used it is common to include all proteins in a dataset; however, it is imperative to consider the biological plausibility of the individual variables acting as part of a unidimensional system/process. On a broader note, it is important to underscore that all analytic decisions be made informed by the available biological knowledge about immunology as a system (e.g., which proteins regulate one another, the timeline of acute reactivity).

In addition to careful consideration of the biological plausibility of this decision, it is imperative that these aggregates are physiometrically evaluated (e.g., reliability, dimensionality; Moriarity and Alloy, 2021; Segerstrom and Smith, 2012) to test their suitability for research. This is particularly important for multifaceted biological processes such as inflammation (Felger and Miller, 2020; Lynall et al., 2020), as discrete inflammatory processes could have different predictors and sequelae of interest. A common example of this possibility is discussion of how pro- vs. anti-inflammatory processes might be differentially associated with psychological outcomes (e.g., depression; Hayley, 2011). Further, the not uncommon practice of creating inflammatory aggregates by standardizing and summing all individual proteins in a dataset can result in composites that are inadvisable for use (Moriarity et al., 2021a) relative to individual

proteins or composites created using standard aggregate creation procedures (e.g., factor analysis). To clarify, this is not a condemnation of inflammatory composites/latent factors; rather, it is an argument that standard data aggregation procedure should be followed to determine if the resulting variables are viable. Consequently, if composites are determined to be biologically-plausible, it is strongly recommended they are empirically investigated before use and measurement properties germane to composite measures (e.g., internal consistency) are reported.

We Got the Power (Analyses)

Psychoneuroimmunology is far from the only field suffering from a dearth of power analyses reported in manuscripts, but it is an important area with room for improvement. Monte Carlo simulations provide a relatively straightforward way to determine power to detect an effect of interest with the sample size available. This technique is also incredibly flexible, able to evaluate the power for any type of analysis. Briefly, this procedure (code provided) involves simulating a dataset with the effect size under question specified and the sample size available to the researchers repeatedly (at least 1,000 times). The proportion of times the effect is significant across the simulated samples indicates the study's power. To illustrate, consider a hypothetical study testing whether IL-6 response to the influenza vaccine is associated with depression controlling for gender and age in a sample of 65 adults. By first simulating multiple datasets with the hypothesized correlations between each variable, it is possible to run the desired analysis in each dataset to determine power (i.e., how many times the hypothesized effect was significant). Alternatively, for readers less comfortable operating in R, there are also Shiny apps available that facilitate the use of Monte Carlo simulations for specific analyses (e.g., pwrSEM (Wang and

Rhemtulla, 2020) and ANOVA_power, for which links are provided at the end of the supplemental code) without the need for writing code.

Conclusion

This primer seeks to highlight potential areas of improvement in the analytic execution of psychoneuroimmunology research. By improving the standard statistical rigor of the field it is possible to conduct more replicable, informative, and clinically-relevant studies to advance the field at a more efficient, consistent, and meaningful pace.

References

Centers for Disease Control and Prevention, 2009. National Health and Nutrition Examination Survey (NHANES) stored biologic specimens: Guidelines for proposals to use samples and proposed cost schedule.

Courvoisier, D.S., Renaud, O., 2010. Robust analysis of the central tendency, simple and multiple regression and ANOVA: a step by step tutorial. Int. J. Psychol. Res. 3, 78–87. https://doi.org/10.21500/20112084.849

Ernst, A.F., Albers, C.J., 2017. Regression assumptions in clinical psychology research practice-a systematic review of common misconceptions. PeerJ e3323. https://doi.org/10.7717/peerj.3323

Felger, J.C., Miller, A.H., 2020. Identifying Immunophenotypes of Inflammation in Depression: Dismantling the Monolith. Biol. Psychiatry 88, 136–138. https://doi.org/10.1016/j.biopsych.2020.04.024

Hayley, S., 2011. Toward an anti-inflammatory strategy for depression. Front. Behav. Neurosci. 5, 1–7. https://doi.org/10.3389/fnbeh.2011.00019

Horn, S.R., Long, M.M., Nelson, B.W., Allen, N.B., Fisher, P.A., Byrne, M.L., 2018. Replication and reproducibility issues in the relationship between C-reactive protein and depression: A systematic review and focused meta-analysis. Brain Behav. Immun. 73, 85–114. https://doi.org/10.1016/j.bbi.2018.06.016

Kowalski, C.J., 1972. On the Effects of Non-Normality on the Distribution of the Sample Product-Moment Correlation Coefficient. Appl. Stat. 21, 1–12.

https://doi.org/10.2307/2346598

Kuhlman, K.R., Robles, T.F., Dooley, L.N., Boyle, C.C., Haydon, M.D., Bower, J.E., 2018. Within-subject associations between inflammation and features of depression : Using the flu vaccine as a mild inflammatory stimulus. Brain Behav. Immun. 69, 540–547. https://doi.org/10.1016/j.bbi.2018.02.001

Liu, Q., Maxwell, S.E., 2020. Multiplicative treatment effects in randomized pretest-posttest experimental designs. Psychol. Methods 25, 71–87. https://doi.org/10.1037/met0000222

Lynall, M.E., Turner, L., Bhatti, J., Cavanagh, J., de Boer, P., Mondelli, V., Jones, D., Drevets, W.C., Cowen, P., Harrison, N.A., Pariante, C.M., Pointon, L., Clatworthy, M.R., Bullmore, E., 2020. Peripheral Blood Cell–Stratified Subgroups of Inflamed Depression. Biol. Psychiatry 88, 185–196. https://doi.org/10.1016/j.biopsych.2019.11.017

Mac Giollabhui, N., Ellman, L.M., Coe, C.L., Byrne, M.L., Abramson, L.Y., Alloy, L.B., 2020. To exclude or not to exclude: Considerations and recommendations for C-reactive protein values higher than 10 mg/L. Brain. Behav. Immun. https://doi.org/10.1016/j.bbi.2020.01.023

Moriarity, D.P., Alloy, L.B., 2021. Back to basics: The importance of measurement properties in biological psychiatry. Neurosci. Biobehav. Rev. 123, 72–82. https://doi.org/10.1016/j.neubiorev.2021.01.008

Moriarity, D.P., Ellman, L.M., Coe, C.L., Olino, T.M., Alloy, L.B., 2021a. A physiometric investigation of inflammatory composites: Comparison of “a priori” aggregates, empirically-identified factors, and individual proteins. Brain, Behav. Immun. - Heal. 18, 100391. https://doi.org/10.1016/j.bbih.2021.100391

Moriarity, D.P., Horn, S.R., Kautz, M.M., Haslbeck, J.M., Alloy, L.B., 2021b. How handling extreme C-reactive protein (CRP) values and regularization influences CRP and depression criteria associations in network analyses. Brain. Behav. Immun. 91, 393–403. https://doi.org/10.1016/j.bbi.2020.10.020

Moriarity, D.P., Kautz, M.M., Mac Giollabhui, N., Klugman, J., Coe, C.L., Ellman, L.M., Abramson, L.Y., Alloy, L.B., 2020. Bidirectional associations between inflammatory biomarkers and depressive symptoms in adolescents: Potential causal relationships. Clin. Psychol. Sci. 8, 690–703. https://doi.org/10.1017/CBO9781107415324.004

Moriarity, D.P., Mac Giollabhui, N., Ellman, L.M., Klugman, J., Coe, C.L., Abramson, L.Y., Alloy, L.B., 2019. Inflammatory proteins predict change in depressive symptoms in male and female adolescents. Clin. Psychol. Sci. 7, 754–767. https://doi.org/10.1177/2167702619826586

Moriarity, D.P., McArthur, B.A., Ellman, L.M., Coe, C.L., Abramson, L.Y., Alloy, L.B., 2018. Immunocognitive model of depression secondary to anxiety in adolescents. J. Youth Adolesc. 47, 2625–2636. https://doi.org/10.1007/s10964-018-0905-7

Out, D., Hall, R.J., Granger, D.A., Page, G.G., Woods, S.J., 2012. Assessing salivary C-reactive protein: Longitudinal associations with systemic inflammation and cardiovascular disease risk in women exposed to intimate partner violence. Brain. Behav. Immun. 26, 543–551. https://doi.org/10.1016/j.bbi.2012.01.019

Segerstrom, S.C., Boggero, I.A., 2020. Expected Estimation Errors in Studies of the Cortisol Awakening Response: A Simulation. Psychosom. Med. 82, 751–756. https://doi.org/10.1097/PSY.0000000000000850

Segerstrom, S.C., Smith, G.T., 2012. Methods, variance, and error in psychoneuroimmunology research: The good, the bad, and the ugly, in: Segerstrom, S.C. (Ed.), Oxford Handbook of Psychoneuroimmunology. Oxford U Press, New York, NY, pp. 421–432.

Shields, G.S., Slavich, G.M., Perlman, G., Klein, D.N., Kotov, R., 2019. The short-term reliability and long-term stability of salivary immune markers. Brain. Behav. Immun. 81, 650–654. https://doi.org/10.1016/j.bbi.2019.06.007

Szabo, Y.Z., Burns, C.M., Lantrip, C., 2022. Understanding associations between rumination and inflammation: A scoping review. Neurosci. Biobehav. Rev. 135, 104523. https://doi.org/10.1016/j.neubiorev.2022.104523

Tabachnick, B.G., Fidell, L.S., 2013. Using multivariate statistics, Sixth. ed. Pearson, Boston, MA.

Wang, A.Y., Rhemtulla, M., 2020. Power analysis for parameter estimation in structural equation modeling: a discussion and tutorial. Adv. Methods Pract. Psychol. Sci. 211. https://doi.org/10.3130/aijt.5.211_5