**Pre-Registration Template for Studies Involving Spatial Analyses**

**Peter Kedron**

School of Geographical Sciences and Urban Planning

Spatial Analysis Research Center (SPARC)

Arizona State University

Tempe, AZ 85281

**Joseph Holler**

Department of Geography

Middlebury College

Middlebury, VT 05753

**Zach Hilgendorf**

School of Geographical Sciences and Urban Planning

Earth Surface Processes and Geomorphology Lab

Arizona State University

Tempe, AZ 85281

Version 1.0 | Created February 12, 2021 | Last Updated February 15, 2021

**Abstract**

Study registration is one way to address *publication bias*, the systematic publication of results based on criteria preference for certain study outcomes; and *specification searching*, the selective reporting of only some analyses conducted during the course of a study. We present a study pre-registration template for studies involving spatial analyses. Our template is meant as a supplement to widely used study pre-registration templates such as those developed by the Open Science Framework (OSF), AsPredicted, and the Berkeley Initiative for Transparency in Social Sciences. We developed this template by reviewing existing templates, integrating their shared components, and then adding new components needed to capture forms of analytical flexibility specific to spatial data analysis. The OSF pre-registration is our most direct source and some of the language included here is intentionally a direct reproduction of that template. The authors of the original template are due full credit for that work. Throughout this document, our additions are those characteristically spatial considerations highlighted in red.

This template is meant to be flexible enough to be adapted to many different forms of spatial analysis. As such, it does not cover every specific aspect of all possible analyses. At the same time, this document is meant as a guide. We recognize that in some instances it may not be possible to provide selected geographic information due for instance to ethical considerations. In such cases, a statement should be made explaining the reasoning for the omission, ideally with reference to any relevant institutional review documentation.

**Keywords:** Pre-analysis Plans, Study Registration, Spatial Analysis, Human Environment and Geographical Sciences

**Pre-Registration Template for Studied Involving Spatial Analyses**

**Study Information**

1. Provide the working title of the study
2. Name the authors of the pre-registration
3. Give an abstract length description of the study. Include background, purpose, overall research question, and the spatial extent of the study area and scale(s) of analysis.
4. List research questions and hypotheses. For each research question provide one or more testable hypotheses. If a hypothesis is directional, state the direction. List interaction or moderation as separate hypotheses.
   1. Identify the spatial coverage over which each hypothesis is expected to hold, and the spatial coverage at which each hypothesis will be tested (e.g., the entire study area, a specific sub-region)
5. Identify where any code or data created or used during the course of the study will be stored and/or made available[[1]](#endnote-1).
6. Define the computational environment that will be used in the analysis when appropriate[[2]](#endnote-2).
   1. Include information about both the hardware (e.g., processor) and software (e.g., OS, version number) environments.

**Study Design**

1. Identify the type of study (e.g., experiment, observational, meta-analysis).
   1. For experimental studies, describe if binding will be used and who will be aware of experimental manipulations.
2. Describe the design of the study (e.g., cross-sectional spatial regression analysis of secondary data). Be as detailed as is necessary given the specifics of their design.
3. Identify whether this study will be confirmatory, evaluating specific hypothesis; or exploratory, without specific hypothesis.
   1. For exploratory analyses, please identify which variables will be explored, and how.
   2. If you intend to conduct confirmatory tests for any hypothesis generated during the exploratory analyses, describe how you will separate those tests from the exploratory analyses
4. Identify the geographic extent of the study.
   1. Provide a written description of the study location and extent. Whenever possible, provide specific geographic coordinates bounding the study extent or a spatial reference file.[[3]](#endnote-3)
5. Describe any randomization that will be used in the study.
   1. Identify any spatial components of the randomizations (e.g., stratified randomization where strata are defined by location)

**Sampling Plan**

1. Identify whether registration will take place before seeing the data; or after seeing the data. Sample, descriptions presented by the OSF pre-registration template include:
   1. Registration prior to creation of data: As of the date of submission of this research plan for preregistration, the data have not yet been collected, created, or realized.
   2. Registration prior to any human observation of the data: As of the date of submission, the data exist but have not yet been quantified, constructed, observed, or reported by anyone - including individuals that are not associated with the proposed study. Examples include museum specimens that have not been measured and data that have been collected by non-human collectors and are inaccessible.
   3. Registration prior to accessing the data: As of the date of submission, the data exist, but have not been accessed by you or your collaborators. Commonly, this includes data that has been collected by another researcher or institution.
   4. Registration prior to analysis of the data: As of the date of submission, the data exist and you have accessed it, though no analysis has been conducted related to the research plan (including calculation of summary statistics). A common situation for this scenario when a large dataset exists that is used for many different studies over time, or when a data set is randomly split into a sample for exploratory analyses, and the other section of data is reserved for later confirmatory data analysis.
   5. Registration following analysis of the data: As of the date of submission, you have accessed and analyzed some of the data relevant to the research plan. This includes preliminary analysis of variables, calculation of descriptive statistics, and observation of data distributions. Please see cos.io/prereg for more information.
2. Explain if/how you have taken steps to not explore and analyze the data (Not needed for exploratory research).
3. Describe how data will be collected using a set of instructions that another person could use to repeat the data collection procedure and recreate the study population.
   1. For human subjects research, include the population from which you obtain subjects, recruitment efforts, payment for participation, how subjects will be selected for eligibility from the initial pool (e.g. inclusion and exclusion rules), and your study timeline.
   2. For studies that do not include human subjects, include information about how you will collect samples, duration of data gathering efforts, source or location of samples
   3. For studies in which the location of subjects or samples must be preserved, explain the motivation for the obfuscation and identify the type of geographic masking technique that will be used[[4]](#endnote-4).
   4. If specific circumstances related to the location or geographic context of the study make it unlikely that the study population could be recreated, present the methods for creating this unique population and give the rational for studying this unique set of subjects.
4. Describe the spatial sampling design that will be used during data collection (e.g., spatially stratified random sampling, systematic spatially unaligned sampling).
   1. Include a description of the size of the sample that also presents how many observations will be collected in different geographic strata or levels if using a stratified, clustered, or multilevel design.
   2. Provide a rationale for the sample size (e.g., funding constraints)
   3. Identify any geographic concerns that will or may constrain sample collection (e.g., in accessibility of sub-regions within the study area).
   4. If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.

**Data Description**

1. Briefly describe the dataset(s), and any sub-set(s) of the data that will be used in this study.
   1. If selected datasets or sub-sets cover only portions of the overall study area, clearly identify which datasets are associated with which locations.
2. Identify if the data is open or available to the public
3. Identify how the data can be accessed by providing: a persistent identifier (DOI), weblink, or description of how to otherwise obtain the dataset
4. Provide the data access/download date for each author
   1. Example: PJK Downloaded 09 February 2021, Accessed 15 February 2021
   2. Identify which dataset will be used in analysis (e.g., Data accessed by PJK on 15 February 2021 will be used for all analyses.
5. Describe how the original data were collected. If the original data collection procedure is well-documented, provide a link to that information.
   1. Attempt to convey the representativeness of the sample and any potential biases.
   2. Identify the spatial data model that will be used in the analysis (e.g. vector, raster)
6. Provide available documentation of or links to metadata or codebooks for the dataset.
7. Describe any prior knowledge you have of the data that is relevant to the analyses in this study.
   1. For each author provide identifiers for prior studies that used this data and details about the variables they previously analyzed.
   2. For authors who have worked on the topic of the study in the geographic location being investigated, briefly summarize each author’s experiences in that location

**Variables**

1. Identify any experimentally manipulated variables (e.g., engineered interventions in stream systems) generated during the creation of the original data that will be used in this study. Provide details about if/how these variables were adapted for use in this study.
2. Identify any measured variables that will be used in this study.
   1. Identify and describe both the response variable(s) and the predictor variable(s)
3. Describe any variable transformations (e.g., log-scaled, categorical) that will occur as part of the study.

**Analyses**

*Considerations for Geographical Analysis:*

1. Identify the coordinate system(s) and projection(s) that will be used during analysis
2. Identify and describe the spatial scale(s) at which the analysis will be conducted
   1. If known, present the characteristic scale of the phenomena being examined
   2. Identify the spatial coverage that will be used when testing each hypothesis
   3. Identify the spatial support (spatial resolution, unit of analysis) that will be used when testing each hypothesis
3. Specify if/how edge effects will be addressed in the analysis. Provide documentation of the extent of any buffer areas or guard areas used.
4. Describe how all spatial sub-groups used in any analysis will be created.
5. Describe any changes that will be made to variables to account for expected first-order spatial effects (e.g., sub-regional differences in means) prior to analysis.
6. Describe how second-order spatial effects (e.g., spatial dependencies) will be measured and adjusted for during the analyses.
7. Identify any spatial anisotropies (directional trends) expected to be observed in the data and describe how any adjustment will be made to the analyses to account for these effects.

*Data Exclusion and Adjustment:*

1. Describe how any type of planned spatial aggregation/disaggregation/rescaling/interpolation will occur for each variable.
2. Identify any data that will be excluded from the analyses and why.
   1. Provide a clear description of the exclusion rule or criteria
   2. Report the original sample size and the sample size after the exclusion rule is applied
   3. Report the location of data excluded from the study (e.g., is exclusion likely to reduce/eliminate coverage in a particular sub-region)
3. Explain how the analyses will account for missing data and provide a detailed description of how the plan to handle missing data will be implemented.
   1. When possible, provide a description of the quantity of missing data for each variable
   2. If spatial interpolations will be made, provide a description of the interpolation procedure that will be applied to each variable
4. Explain how the analyses will handle the presence of outliers.
   1. When possible, estimate the number of outliers expected for each variable[[5]](#endnote-5).
   2. When possible, explain how the analysis will assess and handle the location of outliers.
5. Identify whether any sample weighting will be used in the study, and describe how and why that weighting scheme will be implemented.
   1. Separately identify any spatial component used in the weighting scheme.

*Analytical Specification:*

1. Describe the exact analytical specification that will be used for each hypothesis test.
   1. Include the type of model, the specification of that model, distributional assumptions of the model, and any post-hoc analyses
   2. If possible, specify a predicted or minimal effect size of interest for the relevant variables
   3. If possible, present a statistical power analysis to detect the effect size for each hypothesis
2. Describe all applicable analytical specifications that are spatial in nature. Some common examples:
   1. If a spatial weighting scheme will be used, provide a functional description of that scheme
   2. If a spatial multi-level model is used, identify the spatial scale of each level, the variable included at each level, and the levels any spatial structures or cross-scale structure are estimated at.

*Inference Criteria and Robustness:*

1. Describe the specific criteria (e.g., *p-*values, effect size, model fit) and thresholds that will be used to make inferences
   1. If an adjustment for multiple testing (e.g., Bonferroni, Sidak) will be used, identify the type of adjustment and specify how it will be implemented
2. Explain how the strength, reliability, and robustness of inferences about the focal hypothesis of the study will be tested.
   1. Identify each spatial component that will be varied during robustness or sensitivity analyses

*Exploratory Analyses and Contingency Planning:*

1. Present a brief contingency plan that will be followed should an unanticipated analytical problem arise during the study.
2. Describe any exploratory analyses that will occur and how the outcomes of those exploratory analyses may lead to changes in the confirmatory analyses presented elsewhere[[6]](#endnote-6).

**Integrity Statement**

1. Include an integrity statement - The authors of this preregistration state that they filled out this preregistration to the best of their knowledge and that no other preregistration exists pertaining to the same hypotheses and dataset.
2. Include a time-stamp of the pre-analysis plan and ideally pre-register the plan on a trusted registry (e.g., OSF).

**References**

*Existing Templates Reviewed and Used to Develop This Template:*

AsPredicted. 2021. “Registration Template” Wharton School of Business. Accessed at: <https://osf.io/fnsb6/>

Aust, F. 2020. “prereg: R Markdown Templates for Preregistrations of Scientific Studies” GitHub Repository, <https://github.com/crsh/prereg>

Brandt, Mark J., Hans IJzerman, Ap Dijksterhuis, Frank J. Farach, Jason Geller, Roger Giner-Sorolla, James A. Grange, Marco Perugini, Jeffrey R. Spies, and Anna Van't Veer. "The replication recipe: What makes for a convincing replication?." *Journal of Experimental Social Psychology* 50 (2014): 217-224.

Mellor, David T, and Alexander C DeHaven. 2020. “Templates of OSF Registration Forms.” OSF. November 9. osf.io/zab38.

Mellor, David T, Jolene Esposito, Tom E Hardwicke, Brian A Nosek, Johanna Cohoon, Courtney K Soderberg, Mallory Kidwell, et al. 2019. “Preregistration Challenge: Plan, Test, Discover.” OSF. September 19. osf.io/x5w7h.

Van den Akker, Olmo, Sara J. Weston, Lorne Campbell, William J. Chopik, Rodica I. Damian, Pamela Davis-Kean, Andrew N. Hall, et al. 2019. “Preregistration of Secondary Data Analysis: A Template and Tutorial.” PsyArXiv. November 20. doi:10.31234/osf.io/hvfmr.

van ’t Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2–12. doi: [10.1016/j.jesp.2016.03.004](https://doi.org/10.1016/j.jesp.2016.03.004)

*Selections from the Relevant Literature:*

Burlig, Fiona. "Improving transparency in observational social science research: A pre-analysis plan approach." *Economics Letters* 168 (2018): 56-60.

Coffman, Lucas C., and Muriel Niederle. "Pre-analysis plans have limited upside, especially where replications are feasible." *Journal of Economic Perspectives* 29, no. 3 (2015): 81-98.

Ledgerwood, Alison. "The preregistration revolution needs to distinguish between predictions and analyses." *Proceedings of the National Academy of Sciences* 115, no. 45 (2018): E10516-E10517.

Mertens, Gaëtan, and Angelos-Miltiadis Krypotos. "Preregistration of analyses of preexisting data." *Psychologica Belgica* 59, no. 1 (2019): 338.

Miguel, Edward, Colin Camerer, Katherine Casey, Joshua Cohen, Kevin M. Esterling, Alan Gerber, Rachel Glennerster et al. "Promoting transparency in social science research." *Science* 343, no. 6166 (2014): 30-31.

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. "The preregistration revolution." *Proceedings of the National Academy of Sciences* 115, no. 11 (2018): 2600-2606.

Nust, D., Pebesma, E. (2020). “Practical reproducibility in geography and geosciences.” *Annals of the American Association of Geographers* <https://doi.org/10.1080/24694452.2020.1806028>

Ofosu, George, and Daniel Posner. "Pre-analysis plans: A stocktaking." (2019).

Olken, Benjamin A. "Promises and perils of pre-analysis plans." *Journal of Economic Perspectives* 29, no. 3 (2015): 61-80.

1. Best practices are, whenever possible, to store the code and data in a searchable, open repository (e.g., OSF, FigShare) that will assign these objects a persistent identifier. For computational workflow it is also helpful to provide information about the computational environment in which any analysis was conducted. All these elements can also be containerized to ease reproduction. [↑](#endnote-ref-1)
2. Nust and Pebesma (2020) suggest a combination of literate programming and containerization of final work products can greatly facilitate computational reproducibility by capturing information about hardware and software environments. If containerization will be used it would similarly be useful to note in the preregistration. [↑](#endnote-ref-2)
3. Specific coordinates can be obtained from OpenStreetMap.org by searching a feature and downloading its XML notation. [↑](#endnote-ref-3)
4. It may be inappropriate to provide complete details of the geographic masking procedure as this could well defeat the anonymizing purpose of the procedure. Nonetheless, a general description of the procedure will help others account for masking during their assessment of study results. [↑](#endnote-ref-4)
5. For confirmatory analyses, Van Den Akker et al. (2019) suggest masking the data to obtain this estimate to preserve investigator ignorance of data patterns. [↑](#endnote-ref-5)
6. Van Den Akker et al. (2019) suggest clearly differentiating these exploratory analyses from the confirmatory analyses in the final report (e.g., by labeling them in tables and figures) [↑](#endnote-ref-6)