Pre-Analysis Plan (PAP) Guide*

Danilo Freire†

July 01, 2024

Abstract

This document presents a short guide for creating pre-analysis plans (PAPs) in social and biological sciences, incorporating best practices and guidelines from leading researchers and institutions. The template includes detailed sections on formulating research questions and hypotheses, specifying sample characteristics and data collection procedures, defining outcome measures, outlining statistical models and analyses, and addressing challenges such as multiple hypothesis testing. The template also includes instructions for using Quarto, a free and open-source publishing system, to create reproducible and transparent research documents.

Keywords: Markdown, Pre-Analysis Plan, Quarto, R, Reproducibility, Transparency

*Please refer to the Acknowledgement section for the sources of this document.

†Visiting Associate Professor, Department of Quantitative Theory and Methods, Emory University. danilo.freire@emory.edu, https://danilofreire.github.io.

Contents

1 How to Use This Template

This template is written in Markdown and uses the Quarto document format. Quarto is a powerful tool that allows you to write in Markdown and convert your document to a variety of formats, including PDF, HTML, and Word. You can use this template to write your PAP in Markdown and then convert it to a PDF for submission.

For you to use this template, you need to have R, Pandoc and Quarto installed on your computer. You can install Pandoc by following the instructions on the Pandoc website. To install the Quarto package, run the following command in R:

```
install.packages(c("quarto", "rmarkdown", "tidyverse"))
```

You also need to install a LaTeX distribution on your computer to convert your Markdown document to a PDF. I recommend you install TinyTeX, a lightweight distribution of LaTeX. You can install TinyTeX by running the following command in R:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

For those who prefer a more comprehensive TeX distribution, you can install MiKTeX on Windows or MacTeX on macOS.

Finally, you should also install the Libertine and Inconsolata fonts on your computer to use this template. You can download the Libertine font from the Linux Libertine website and the Inconsolata font from the Google Fonts website. However, you can use any other font you prefer by changing the `fontfamily` and `monofont` options in the YAML header of this document. More information on how to change the font in your document can be found in the Quarto documentation.

You can then use the Quarto package to convert your Markdown document to a PDF. To convert your document to a PDF, run the following command in R:

```
quarto::quarto_render("pre-analysis-plan.qmd", output_format = "pdf")
```

This command will convert your Markdown document to a PDF and save it in the same directory as your Markdown file.

2 How to Write Your Text Using Markdown

Markdown is a lightweight markup language that you can use to format your text. You can use Markdown to create headings, lists, tables, equations, and figures. Here are some examples of how to format your text using Markdown:

2.1 Headings

You can create headings using the # symbol. For example, `# Heading 1` creates a first-level heading, `## Heading 2` creates a second-level heading, and so on. You can create up to six levels of headings using the # symbol.

2.2 Lists

To create an ordered list with nested unordered sub-items in Quarto, you can write the following code:

```
1. This is an ordered list.
2. This is the second item in the ordered list.
  - This is a sub-item in the unordered list.
    - This is a sub-sub-item in the unordered list.
```

1. This is an ordered list.
2. This is the second item in the ordered list.
    - This is a sub-item in the unordered list.
        - This is a sub-sub-item in the unordered list.

You can also create unordered lists:

```
- This is an unordered list.
- This is the second item in the unordered list.
    - This is a sub-item in the unordered list.
```

- This is an unordered list.
- This is the second item in the unordered list.
    - This is a sub-item in the unordered list.

2.3 Tables

You can create tables using the | symbol. For example:

```
Table: Your Caption {#tbl-markdown}

| A            | New               | Table          |
|:-------------|:-----------------:|---------------:|
|left-aligned  |centre-aligned     |right-aligned   |
|*italics*     |~~strikethrough~~  |**boldface**    |
```

Table 1: Your Caption

A	New	Table
left-aligned	centre-aligned	right-aligned
italics	~~strikethrough~~	**boldface**

The : symbols in the second row of the table determine the alignment of the text in each column. You can use `left`, `center`, or `right` to align the text.

To use the `kable` and `kableExtra` packages in `R` for creating and customising tables, you first need to install these packages if you have not already. You can do this by running `install.packages("knitr")` and `install.packages("kableExtra)` in `R`. Below is an example. Please check the `pre-analysis-plan.qmd` file for the code.

```
library(kableExtra)
df <- mtcars[1:5, 1:6]
kable(df, "latex", booktabs = T) %>%
  kable_styling(position = "center", font_size = 12) %>%
  add_header_above(c(" " = 1, "Group 1" = 2, "Group 2" = 2,
                     "Group 3" = 1, "Group 4" = 1)) %>%
  add_header_above(c(" ", "Group 5" = 4, "Group 6" = 2), bold = T) %>%
  footnote(general = "Your comments go here.")
```

Table 2: A table created with kable and kableExtra.

	Group 5				**Group 6**	
	Group 1		Group 2		Group 3	Group 4
	mpg	cyl	disp	hp	drat	wt
Mazda RX4	21.0	6	160	110	3.90	2.620
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875
Datsun 710	22.8	4	108	93	3.85	2.320
Hornet 4 Drive	21.4	6	258	110	3.08	3.215
Hornet Sportabout	18.7	8	360	175	3.15	3.440

Note:
Your comments go here.

This will create a table with the `mtcars` dataset from the `datasets` package in `R`. You can customise the table using the `kableExtra` package to add a caption, change the font size, and add headers above the columns. You can also add footnotes to the table using the `footnote` function.

2.4 Equations

You can create equations using the `$$` symbol. For example in Equation 1, we have the formula for the standard deviation of a population:

```
$$
\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}}
$$ {#eq-stddev}
```

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}} \tag{1}$$

You can also create equations inline by using the `$` symbol. For example, `$\alpha = \beta + \gamma$` will render as $\alpha = \beta + \gamma$. To learn more about how to write equations in LaTeX using Markdown, you can refer to the Overleaf documentation.

2.5 Figures

You can include figures in your document using the `![Caption](path/to/image.png){#fig-label}` syntax. For example:

```
![This is a figure caption.](path/to/image.png){#fig-label}
```

This will include the image `path/to/image.png` in your document with the caption "This is a figure caption." You can refer to the figure using the label `fig-label`.

You can also use the `knitr` package in `R` to create figures. This gives you more control over the appearance of the figure and allows you to include the code used to create the figure in your document. To include the same figure, you can use the following code:

```
library(knitr)
knitr::include_graphics("path/to/image.png")
```

There are many options you can use to customise the appearance of your figures, such as changing the width of the figure, adding a border, or changing the alignment. You can find more information on how to customise figures in the Quarto documentation.

2.6 Code Blocks and Syntax Highlighting

You can include code blocks in your document using triple backticks (```). For example, if you would like to include an `R` code block in your document, you can write ```` ```{r} ````, and close it with another set of triple backticks (```). You can also add chunk options to the code block to control the output of the code block with #|. For example:

```
#| label: fig-airquality
#| fig-cap: "Temperature and ozone level."
#| warning: false

library(ggplot2)
ggplot(airquality, aes(Temp, Ozone)) +
  geom_point() +
  geom_smooth(method = "loess", se = FALSE)
```

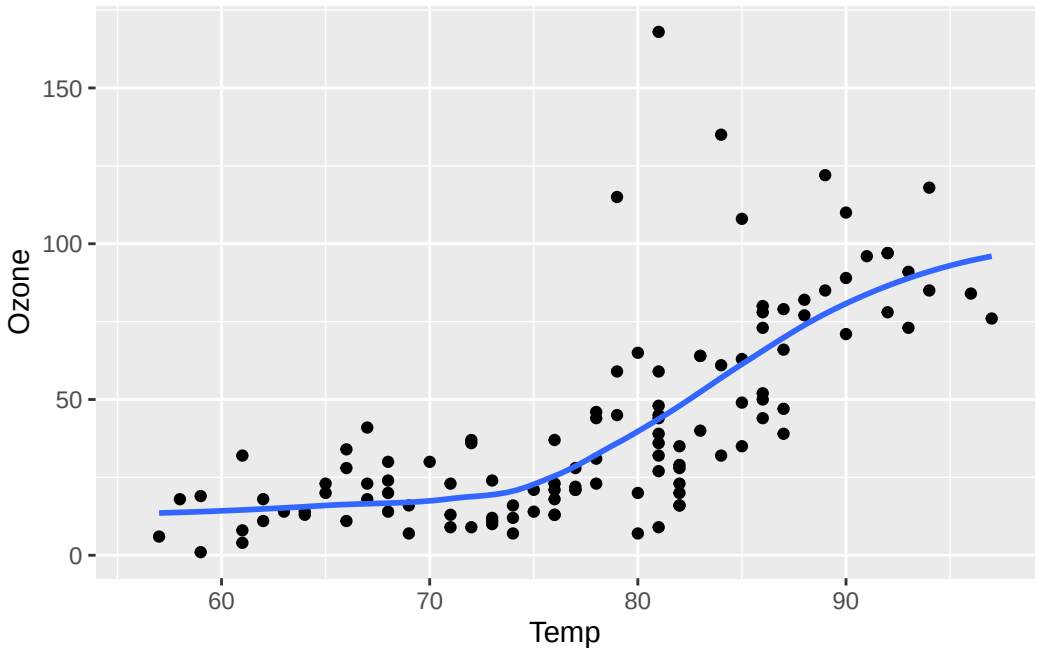

Figure 1: Temperature and ozone level.

Quarto will run the code in the code block and include the output in the final document, as you can see in Figure 1. To learn more about code cell options, you can refer to the Quarto documentation.

2.7 Citations

You can include citations in your document using the `@` symbol followed by the citation key. For example, `@freire2018evaluating` will create a citation to the reference with the key `freire2018evaluating`. You can also include page numbers in your citation by writing `@freire2018evaluating [10--15]`. The result will be: Freire (2018, 10–15). To include multiple citations in one bracket, you can write `[@freire2018evaluating 10; @mignozzetti2022legislature 15]`. This will create a citation like this: (Freire 2018, 10; Mignozzetti et al. 2022, 15).

To include a bibliography in your document, you need to create a `.bib` file with your references and include it in the YAML header of your document. As you can find in the YAML header of this document:

```yaml
bibliography: references.bib
```

When searching for academic references, Google Scholar is an excellent tool for finding BibTeX citations. Here is a quick guide:

1. Search for your desired reference on Google Scholar
2. Look for the "Cite" button below the reference
3. Click on "Cite" and then select "BibTeX" from the options
4. Copy the generated BibTeX citation
5. Paste it into your `.bib` file

This method makes it easy to maintain consistent and accurate citations in your academic work. Plus, it saves you the hassle of manually formatting each reference.

2.8 Footnotes

Quarto supports inline footnotes[1]. To create a footnote, simply add a caret and brackets with a label inside, like this: `[^label]`. Then, you can define the footnote content anywhere in the document using the same label followed by a colon[2]. I usually include them at the end of a paragraph. Quarto will automatically number and format your footnotes for you.

[1] This is an inline footnote.

[2] You can also include multiple paragraphs in a footnote by indenting the subsequent paragraphs.

3 A Guide to Pre-Analysis Plans (PAPs)

Pre-Analysis Plans (PAPs) are detailed documents that researchers create before collecting or analysing data, outlining their intended research methodology, hypotheses, and analytical approach. These plans have become increasingly important in social and biological sciences as a tool to enhance research transparency, credibility, and reproducibility. PAPs serve as a roadmap for researchers, helping them think through their study design and analysis strategy in advance, while also providing a public record of their intentions.

The use of PAPs has grown significantly in recent years, particularly in fields such as economics, political science, and psychology. According to the American Economic Association's website, as of 2024, the AEA RCT Registry currently lists 8,937 studies with locations in 169 countries[3]. This trend reflects a growing recognition of the importance of research transparency and the need to address issues such as publication bias and p-hacking.

4 Purpose and Benefits of PAPs

4.1 Enhancing Research Transparency

PAPs provide a clear record of the researcher's intentions before data collection or analysis, reducing concerns about data mining or selective reporting. By specifying analyses in advance, researchers demonstrate their commitment to transparent and unbiased research practices. This transparency allows other researchers and stakeholders to distinguish between pre-planned analyses and exploratory findings, increasing confidence in the reported results.

For example, in a study on the effects of a cash transfer program on educational outcomes, researchers might specify in their PAP that they will analyse the impact on primary school enrollment rates, test scores, and dropout rates. By doing so, they commit to reporting these outcomes regardless of whether the results are statistically significant or align with their hypotheses. This approach prevents the temptation to selectively report only positive or interesting findings after seeing the data.

[3] See: https://www.socialscienceregistry.org/. Access: July 2024.

4.2 Improving Research Quality

By forcing researchers to think through their analysis in advance, PAPs can lead to better-designed studies and more robust analytical approaches. This process often reveals potential issues or oversights that can be addressed before data collection begins. For instance, when drafting a PAP, researchers might realise they need to include additional control variables or consider potential confounding factors they had not initially thought of.

Consider a medical study examining the effectiveness of a new drug. In developing the PAP, researchers might recognise the need to stratify their sample based on age groups or pre-existing conditions, leading to a more nuanced and informative analysis. They might also identify potential interaction effects between the drug and other medications, prompting them to collect data on participants' full medical histories.

4.3 Addressing Publication Bias

PAPs can help combat publication bias by encouraging the reporting of all pre-specified analyses, regardless of their results. This approach ensures that null or unexpected findings are not selectively omitted from publication. Publication bias is a significant issue in many fields, where studies with positive or novel results are more likely to be published than those with null or negative findings.

For example, in psychology, the "file drawer problem" refers to the tendency for studies with non-significant results to remain unpublished. By pre-registering their analyses, researchers commit to reporting all findings, which can help provide a more balanced view of the evidence. In a meta-analysis of psychological interventions, for instance, including pre-registered studies might reveal that the overall effect size is smaller than what would be estimated based solely on published literature.

4.4 Increasing Credibility

Pre-registering analyses can increase the credibility of research findings, as it demonstrates that the results were not cherry-picked or manipulated to achieve desired outcomes. This is particularly important in fields where research findings can influence policy decisions or clinical practices.

For instance, in a study on the effectiveness of a new educational intervention, pre-registering the primary outcome measures (e.g., standardised test scores) and analysis methods prevents researchers from selectively reporting only the most favorable results. If the intervention shows mixed effects – perhaps improving math scores but not reading scores – the PAP ensures that both outcomes are reported, providing a more accurate picture of the intervention's impact.

4.5 Facilitating Replication

Detailed PAPs make it easier for other researchers to replicate studies, as they provide a clear roadmap of the intended analyses and methodologies. Replication is crucial for scientific progress, but it often fails due to insufficient information about the original study's methods.

For example, in a complex econometric analysis of labor market policies, a well-written PAP would specify not only the data sources and variables used but also the exact statistical models, including any transformations or coding of variables. This level of detail allows other researchers to reproduce the analysis, even if they do not have access to the original data, by following the same procedures with similar datasets.

5 Key Components of a PAP

5.1 Research Questions and Hypotheses

Clearly state the primary research questions and hypotheses to be tested. Be specific about what you expect to find and why. This section should provide a theoretical framework for the study and explain how the hypotheses relate to existing literature. For instance:

Research Question: "*Does providing microfinance loans to women in rural areas increase their household income and empowerment?*"

Hypotheses:

1. Women who receive microfinance loans will show a 20% increase in household income after one year compared to the control group.

2. Loan recipients will report higher scores on a standardised empowerment index, with an expected effect size of 0.5 standard deviations.
3. The impact of loans on income will be moderated by the recipient's level of education, with more educated women showing larger income gains.

5.2 Sample and Data Collection

Describe the sample size, sampling method, and data collection procedures in detail. Include information on:

- Target population
- Inclusion and exclusion criteria
- Recruitment methods
- Sample size calculations and power analysis

For instance, in a study on the effectiveness of a new teaching method:

"We will recruit 1000 students from 50 public high schools in urban areas of California. Schools will be randomly selected from a list of all eligible schools (defined as those with at least 500 students and a 30% or higher free lunch program participation rate). Within each school, we will randomly select 20 students from the 10th grade. Power calculations indicate that this sample size will allow us to detect an effect size of 0.2 standard deviations with 80% power at a 5% significance level, accounting for potential clustering at the school level."

5.3 Outcome Measures

Define primary and secondary outcome measures, including how they will be constructed from raw data. Provide detailed information on:

- Variable definitions
- Measurement scales
- Data sources
- Any planned data transformations

Example for a health intervention study: "Primary outcome: Body Mass Index (BMI), calculated as weight in kilograms divided by height in meters squared. Weight will be measured using a calibrated digital scale, and height will be measured using a stadiometer. Both measurements will be taken twice and averaged.

Secondary outcomes:

1. Physical activity level: Measured using the International Physical Activity Questionnaire (IPAQ), scored according to the official IPAQ scoring protocol.
2. Dietary quality: Assessed using a 7-day food diary, which will be analysed to create a Healthy Eating Index (HEI) score ranging from 0-100.
3. Self-efficacy for health behaviors: Measured using the Health-Specific Self-Efficacy Scale, a 10-item Likert scale questionnaire."

5.4 Statistical Models and Analyses

Specify the statistical models and analyses to be used, including:

- Type of statistical tests (e.g., t-tests, ANOVA, regression)
- Model specifications
- Handling of missing data
- Treatment of outliers
- Planned subgroup analyses or robustness checks

For example, in an economic study on the impact of a job training program:

"Main analysis: We will use an OLS regression model to estimate the effect of the job training program on annual income:

$$Income_i = \beta_0 + \beta_1 Treatment_i + \beta_2 BaselineIncome_i + \beta_3 Education_i + \beta_4 Age_i + \beta_5 Gender_i + \varepsilon_i$$

Where $Treatment_i$ is a dummy variable indicating assignment to the job training program.

Missing data: We will use multiple imputation with chained equations (MICE) to handle missing data, creating 20 imputed datasets.

Outliers: Income values above the 99th percentile will be winsorised to reduce the influence of extreme outliers.

Subgroup analysis: We will examine heterogeneous treatment effects by gender and by baseline income quartiles, adding interaction terms to the main regression model.

Robustness checks: We will re-run the main analysis using a difference-in-differences approach and a propensity score matching method to ensure our results are not sensitive to the choice of analytical approach."

5.5 Multiple Hypothesis Testing

Address how multiple hypothesis testing will be handled to control for false positives. Describe methods such as:

- Bonferroni correction
- False Discovery Rate (FDR) control
- Family-wise error rate (FWER) control

Example: "To address multiple hypothesis testing, we will use the following approach:

1. For our three primary outcomes, we will use the Bonferroni correction, adjusting the significance level to $\alpha = 0.05/3 = 0.0167$.
2. For secondary outcomes, we will control the False Discovery Rate using the Benjamini-Hochberg procedure with a q-value of 0.10.
3. For exploratory analyses, we will report unadjusted p-values but clearly label these findings as exploratory and interpret them cautiously."

5.6 Covariates and Control Variables

List all covariates and control variables to be included in the analyses, explaining their relevance to the research questions. For instance:

"We will include the following control variables in our main regression analyses:

1. Age: Continuous variable, measured in years
2. Gender: Binary variable (0 = male, 1 = female)
3. Education level: Categorical variable with four levels (less than high school, high school graduate, some college, college graduate or higher)
4. Baseline income: Continuous variable, measured in dollars
5. Employment status at baseline: Binary variable (0 = unemployed, 1 = employed)

These variables are included because previous research has shown they are strong predictors of labor market outcomes and may influence the effectiveness of job training programs."

5.7 Heterogeneous Effects

If applicable, describe any planned analyses of heterogeneous effects or subgroup analyses. For example:

"We will examine heterogeneous treatment effects for the following subgroups:

1. Gender: We hypothesise that the job training program may have differential effects for men and women due to differences in labor market discrimination and occupational segregation.
2. Age groups: We will analyse treatment effects separately for participants under 30, 30-50, and over 50 years old, as the effectiveness of job training may vary across career stages.
3. Education level: We will examine whether the program's impact differs based on participants' educational attainment.

For each subgroup analysis, we will add interaction terms between the treatment indicator and the subgroup variable to our main regression model."

5.8 Power Calculations

Include power calculations to justify the chosen sample size and demonstrate the ability to detect meaningful effects. For instance:

"We conducted power calculations using `R` and the `DeclareDesign` package. Assuming a two-tailed test with $\alpha = 0.05$ and 80% power, our sample size of 1000 participants (500 per group) allows us to

detect a minimum effect size of $d = 0.18$ for our primary outcome of annual income. This effect size is considered small to medium and is in line with previous studies on similar job training interventions. For our subgroup analyses, the minimum detectable effect size increases to $d = 0.28$, which we consider acceptable for these secondary analyses."

6 When to Use PAPs

PAPs are particularly valuable in:

- Experimental studies, including randomised controlled trials
- Studies with multiple plausible outcome measures
- Research involving subgroup analyses
- Projects with potential for researcher degrees of freedom in analysis
- Large-scale observational studies
- Studies with high-stakes outcomes or policy implications

For example, in a large-scale randomised controlled trial evaluating the impact of a new educational curriculum, a PAP would be crucial. The study might have multiple outcome measures (e.g., test scores, attendance rates, student engagement), various subgroups of interest (e.g., by gender, socioeconomic status, prior achievement), and potential for different analytical approaches. A PAP would help ensure that the primary analyses are clearly defined and distinguished from exploratory analyses, reducing the risk of cherry-picking results or p-hacking.

Similarly, in an observational study examining the long-term health effects of a particular dietary pattern, a PAP would be valuable. The researchers could specify in advance which health outcomes they will focus on, how they will control for potential confounding factors, and what methods they will use to address issues like selection bias. This approach would increase the credibility of the findings, especially given the often conflicting results in nutritional epidemiology.

7 Challenges and Limitations

7.1 Constraining Exploratory Analysis

PAPs may limit researchers' ability to explore unexpected findings in the data. To address this, consider including a section for planned exploratory analyses. For example:

"While our primary analyses will focus on the pre-specified outcomes and methods outlined above, we recognise the potential for unexpected findings. We will include a clearly labeled exploratory analysis section in our final report, where we may investigate:

1. Non-linear relationships between continuous predictors and outcomes
2. Potential mediating factors in the relationship between the intervention and primary outcomes
3. Unexpected patterns or clusters in the data identified through data visualization techniques

Any findings from these exploratory analyses will be clearly labeled as such and interpreted cautiously, with an emphasis on generating hypotheses for future research rather than drawing definitive conclusions."

7.2 Incomplete Specification

It is challenging to anticipate all possible analyses or issues that may arise during the research process. Be as comprehensive as possible while allowing for some flexibility. For instance:

"While we have attempted to be as comprehensive as possible in this PAP, we acknowledge that unforeseen circumstances may necessitate deviations from the plan. Any such deviations will be clearly documented and justified in the final report. Potential scenarios that might require flexibility include:

1. Lower than expected recruitment rates, which may necessitate adjustments to the power calculations and potentially the analytical approach
2. Unexpected data quality issues that may require additional data cleaning or imputation methods
3. External events (e.g., policy changes, economic shocks) that may impact the interpretation of our results

In such cases, we will strive to adhere as closely as possible to the spirit of the original plan while making necessary adjustments to ensure the validity and relevance of our analyses."

7.3 Time and Resource Intensive

Creating a detailed PAP can be time-consuming and may require additional resources. However, this upfront investment often leads to more efficient and robust research in the long run. Researchers should budget time and resources for developing a comprehensive PAP, potentially including:

1. Literature review to inform hypotheses and analytical approaches
2. Consultation with statisticians or methodologists to refine the analytical plan
3. Pilot studies or simulations to inform power calculations and test data collection instruments
4. Team meetings to discuss and refine the PAP
5. External review of the PAP by colleagues or experts in the field

While this process may seem daunting, it can ultimately save time and resources by identifying potential issues early and providing a clear roadmap for the research project.

7.4 Balancing Specificity and Flexibility

Finding the right balance between being specific enough to prevent p-hacking and flexible enough to address unforeseen circumstances can be challenging. One approach is to specify decision rules for potential deviations from the plan. For example:

"We will adhere to the following decision rules for potential deviations from the PAP:

1. If recruitment falls below 80% of the target sample size, we will recalculate power analyses and may adjust our primary outcome to a composite measure to maintain adequate statistical power.
2. If attrition exceeds 20%, we will conduct sensitivity analyses using multiple imputation and inverse probability weighting in addition to our planned complete case analysis.
3. If we discover strong violations of model assumptions (e.g., non-normality, heteroscedasticity), we will employ appropriate transformations or alternative modeling approaches (e.g., generalised linear models, non-parametric tests) and report both the original and alternative analyses."

8 Best Practices for Creating PAPs

8.1 Be Specific but Allow for Flexibility

Provide detailed plans while allowing for some flexibility to address unforeseen circumstances. Consider including decision rules for potential deviations from the plan. For example:

"This PAP outlines our primary analytical strategy, but we acknowledge that unforeseen circumstances may necessitate adjustments. We will adhere to the following principles when considering deviations from the plan:

1. Any deviation must be justified based on methodological grounds or unforeseen data characteristics, not on the results of preliminary analyses.
2. All deviations will be clearly documented and reported in the final manuscript, including the rationale for the change and any implications for the interpretation of results.
3. We will consult with independent experts or our advisory board before implementing major changes to the analysis plan.
4. If substantial deviations are required, we will consider publishing an addendum to the original PAP, clearly stating the changes and their justifications."

8.2 Register Your PAP

Register your PAP with a recognised registry, such as:

- Open Science Framework (OSF)
- EGAP Pre-Analysis Plan Registry
- AEA RCT Registry
- ClinicalTrials.gov (for medical research)

Registration provides a time-stamped, publicly accessible version of your plan, enhancing transparency and credibility.

8.3 Update as Necessary

If changes to the plan are required, document and justify these changes transparently. Consider using version control to track modifications. For example:

"Any updates to this PAP will be logged in a change document, which will include:

- Date of the change
- Description of the modification
- Rationale for the change
- Potential impact on the interpretation of results

This change log will be made publicly available alongside the original PAP."

8.4 Use Clear and Precise Language

Write your PAP in clear, unambiguous language to ensure that other researchers can understand and potentially replicate your planned analyses. Avoid jargon where possible, and define technical terms when they are necessary.

8.5 Include Code and Syntax

When possible, include example code or syntax for planned analyses to further enhance reproducibility. This could be done by:

- Providing annotated code snippets in an appendix
- Linking to a public repository (e.g., GitHub) with full analysis scripts
- Including pseudo-code for complex analytical procedures

8.6 Seek Feedback

Share your PAP with colleagues or mentors for feedback before finalising and registering it. Consider:

- Organising a pre-registration workshop with your research team
- Requesting review from methodologists or subject matter experts
- Presenting your PAP at a departmental seminar for broader input

9 Examples

For examples of well-crafted PAPs, consider reviewing the following resources:

- Open Science Framework (OSF)
- EGAP Pre-Analysis Plan Registry
- AEA RCT Registry
- ClinicalTrials.gov

Feel free to search for PAPs in your field of interest to see how other researchers have approached pre-registration.

10 Acknowledgements

This guide was based on the following websites:

- https://egap.org/resource/10-things-to-know-about-pre-analysis-plans/
- https://www.povertyactionlab.org/resource/pre-analysis-plans
- https://www.bitss.org/resource-tag/pre-analysis-plans
- https://blogs.worldbank.org/en/impactevaluations/how-good-are-pre-analysis-plans-practice-an

Code examples were partially provided by GitHub Copilot, and Perplexity AI was used for text generation. I verified all the information and examples provided in this document.

References

Freire, D. (2018). Evaluating the effect of homicide prevention strategies in São Paulo, Brazil: A synthetic control approach. *Latin American Research Review*, 53(2):231–249.

Mignozzetti, U., Cepaluni, G., and Freire, D. (2022). Legislature Size and Welfare: Evidence from Brazil. *American Journal of Political Science*, pages 1–17.